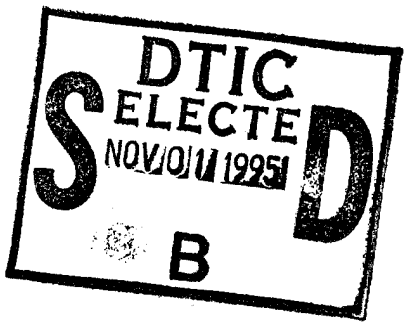
FINAL REPORT

Multiple Target Tracking: Fast Algorithms for Data Association and State Estimation

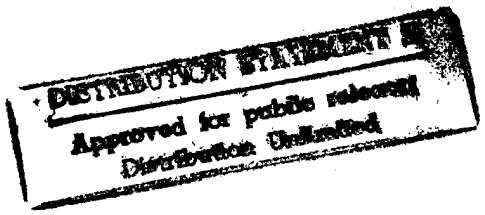
ONR Grant No. N00014-92-J-1755

1 July 1992 - 31 December 1994

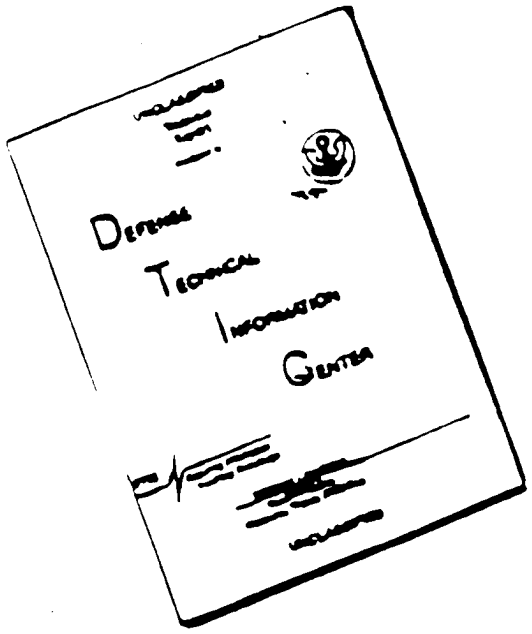


P.I.: Dr. N. K. Bose

Program Manager: Dr. R. N. Madan

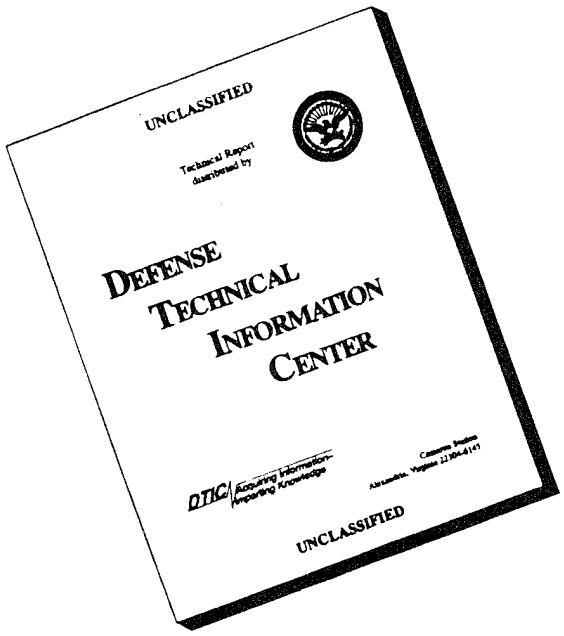


DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

OFFICE OF THE UNDER SECRETARY OF DEFENSE (ACQUISITION)
DEFENSE TECHNICAL INFORMATION CENTER
CAMERON STATION
ALEXANDRIA, VIRGINIA 22304-6145

IN REPLY REFER TO DTIC-OCC

SUBJECT: Distribution Statements on Technical Documents

TO: OFFICE OF NAVAL RESEARCH
CORPORATE PROGRAMS DIVISION
ONR 353
800 NORTH QUINCY STREET
ARLINGTON, VA 22217-5660

1995 1031 021

1. Reference: DoD Directive 5230.24, Distribution Statements on Technical Documents, 18 Mar 87.

2. The Defense Technical Information Center received the enclosed report (referenced below) which is not marked in accordance with the above reference.

FINAL REPORT
N00014-92-J-1755
TITLE: MULTIPLE TARGET
TRACKING: FAST ALGORITHMS
FOR DATA ASSOCIATION AND
STATE ESTIMATION

3. We request the appropriate distribution statement be assigned and the report returned to DTIC within 5 working days.

4. Approved distribution statements are listed on the reverse of this letter. If you have any questions regarding these statements, call DTIC's Cataloging Branch, (703) 274-6837.

FOR THE ADMINISTRATOR:

1 Encl

GOPALAKRISHNAN NAIR
Chief, Cataloging Branch

FL-171
Jul 93

DISTRIBUTION STATEMENT A:

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

DISTRIBUTION STATEMENT B:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; (Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT C:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS; (Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT D:

DISTRIBUTION AUTHORIZED TO DOD AND U.S. DOD CONTRACTORS ONLY; (Indicate Reason and Date Below). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT E:

DISTRIBUTION AUTHORIZED TO DOD COMPONENTS ONLY; (Indicate Reason and Date Below). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT F:

FURTHER DISSEMINATION ONLY AS DIRECTED BY (Indicate Controlling DoD Office and Date Below) or HIGHER DOD AUTHORITY.

DISTRIBUTION STATEMENT X:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE WITH DOD DIRECTIVE 5230.25, WITHHOLDING OF UNCLASSIFIED TECHNICAL DATA FROM PUBLIC DISCLOSURE, 6 Nov 1984 (Indicate date of determination). CONTROLLING DOD OFFICE IS (Indicate Controlling DoD Office).

The cited documents has been reviewed by competent authority and the following distribution statement is hereby authorized.

A

(Statement)

OFFICE OF NAVAL RESEARCH
CORPORATE PROGRAMS DIVISION
ONR 353
800 NORTH QUINCY STREET
ARLINGTON, VA 22217-5660

(Controlling DoD Office Name)

(Reason)

(Controlling DoD Office Address, City, State, Zip)

DEBRA T. HUGHES
DEPUTY DIRECTOR
CORPORATE PROGRAMS OFFICE

19 SEP 1995

(Signature & Typed Name)

(Assigning Office)

(Date Statement Assigned)

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	21 Feb. 95	Final 1 July 92 - 31 December 94

4. TITLE AND SUBTITLE
Multiple Target Tracking: Fast Algorithm for Data Association and State Estimation

5. FUNDING NUMBERS
Grant No.:
N00014-92-J-1755

Modification No.:
A00002

6. AUTHOR(S)
Dr. N. K. Bose
HRB-Systems Professor of Electrical Engineering
Director of Spatial & Temporal Signal Processing Center

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802

8. PERFORMING ORGANIZATION REPORT NUMBER
PSUONR 1755

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Mr. Gerald Smith
Office of Naval Research Regional Director
Federal Building - Second Floor, Room 208
536 South Clark Street
Chicago, IL 60605-1588

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Scientific Program Officer: Dr. R. N. Madan
Office of Naval Research, Code: 1114SE
800 North Quincy Street, Arlington, VA 22217-5000

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Scientific Officer Code: 1114SE (3 copies)
Administrative Grants Officer, ONR (1 copy)
Director, Naval Research Laboratory (1 copy)
Defense Technical Information Center (1 copy)

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

A unified framework is proposed for providing a systematic scheme for generating the data association hypotheses efficiently in the target-oriented, measurement-oriented, and track-oriented approaches to multitarget tracking. A fast recursive algorithm for computing the a posteriori probabilities, suitable for implementation in a distributed multiprocessor system is developed and its links to the theory of permanents is established. An analysis of this algorithm reveals its superiority over existing ones in the average case. In the related problem of direction-of-arrival estimation, a new non-search-type subspace method, called the PESS method, is proposed. This method exploits the structure of the steering matrix more thoroughly to yield a residual-error theoretically shown to be either less than or equal to that obtained by LS-ESPRIT. Furthermore, simulation conducted on several sets of data showed that the PESS method outperforms the TLS-ESPRIT method. Constraints for forcing all roots of a polynomial to the unit circle are obtained for more reliable estimation especially in the low SNR case. Finally, for improved preprocessing to facilitate tracking, a theoretical analysis is proposed to evaluate the robustness of a TLS algorithm, developed earlier, for image reconstruction from a sequence of undersampled noisy and blurred frames.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Multitarget tracking, data association, parameter estimation, sequence processing, fast algorithms, parallel implementation	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Grant Title: Multiple Target Tracking: Fast Algorithms for Data Association and State Estimation

Grant Number: Office of Naval Research Grant N00014-92-J-1755

Grant Duration: Originally from July 1, 1992 - June 30, 1994
Extended to December 31, 1994 by Authority: 10 USC 2358 as amended, and 31 USC 6304

Grant Amount: $120,000

Principal Investigator: Dr. N. K. Bose
HRB-Systems Professor of Electrical Engineering
The Spatial and Temporal Signal Processing Center
Department of Electrical Engineering
The Pennsylvania State University
University Park, PA 16803

Program Manager: Dr. Rabinder N. Madan
1114 SE, Code 414
Department of the Navy
Office of the Chief of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Administrative Grants Officer: Mr. Gerald Smith
Office of Naval Research Regional Director
Federal Building - Second Floor
Room 208
536 South Clark Street
Chicago, IL 60605-1588

CONTENTS

1. Research Abstract

A unified framework is proposed for providing a systematic scheme for generating the data association hypotheses efficiently in the target-oriented, measurement-oriented, and track-oriented approaches to multitarget tracking. A fast recursive algorithm for computing the *a posteriori* probabilities, suitable for implementation in a distributed multiprocessor system is developed and its links to the theory of permanents is established. An analysis of this algorithm reveals its superiority over existing ones in the average case. In the related problem of direction-of-arrival estimation, a new non-search-type subspace method, called the PESS method, is proposed. This method exploits the structure of the steering matrix more thoroughly to yield a residual-error theoretically shown to be either less than or equal to that obtained by LS-ESPRIT. Furthermore, simulation conducted on several sets of data showed that the PESS method outperforms the TLS-ESPRIT method. Constraints for forcing all roots of a polynomial to the unit circle are obtained for more reliable estimation especially in the low SNR case. Finally, for improved preprocessing to facilitate tracking, a theoretical analysis is proposed to evaluate the robustness of a TLS algorithm, developed earlier, for image reconstruction from a sequence of undersampled noisy and blurred frames.

2. Research Description

(a) Background

To solve the problem of data association between targets and measurements, three typical approaches have been reported in the literature. The first is called a *target-oriented* approach in which each measurement is assumed to have originated from either a known target or clutter, as in the *Joint Probabilistic Data Association Filter* (JPDAF) [1]. The scond is called a *measurement-oriented* approach in which each measurement is hypothesized to have originated from either a known target, a new target, or clutter [2]. The third approach is referred to as *track-oriented* where each track is hypothesized to be either undetected, terminated, associated with a measurement, or linked to the start of a maneuver [3].

In recent years, a lot of attention has been given to the task of improving the computational efficiency of a multitarget tracking algorithm. Fitzgerald [4] proposed an *ad hoc* formula to approximately compute the β_j^t's in the JPDAF. For $j > 0$, β_j^t is the *a posteriori* probability that measurement j originated from target t. Fitzgerald's formula breaks down when four targets are required to be tracked [5]. Sengupta and Iltis [5] developed an analog neural network to emulate the JPDAF. They showed that the neural network is capable of handling two to six targets and three to twenty measurements. However, the heuristic nature of their approach makes implementation difficult [6]. Alternatively, Nagarajan, *et al*, [7] arranged the hypotheses in the measurement-oriented approach [2] in a special order so that the probabilities of the hypotheses are proportional to the product of certain probability factors already evaluated. The algorithm locates the N globally best hypotheses without evaluating all of them. In [8], Fisher and Casasent developed a fast joint probabilistic data association (JPDA) algorithm. In their algorithm, the computation of the β_j^t's was implemented using enormous number of vector inner product operations. Since the vector inner product operation could be easily realized on an optical processor, they proposed a specialized optical processor to approximately implement the fast JPDA algorithm. Recently, Zhou and Bose [9] proposed a depth-first search (DFS) approach to efficiently compute the β_j^t's in the JPDAF. Their algorithm requires much less computation than the fast JPDA algorithm in the average case, when there are, on average, two to three measurements inside the validation gate of a target. Although this DFS algorithm is much more efficient than the fast JPDA algorithm in the average case, it is specifically directed towards implementation through a system with a powerful centralized processor, normally used in ground-based tracking systems.

(b) Research Contributions

In this research, an efficient algorithm has been developed to compute the *a posteriori* probabilities of the origins of measurements in the joint probabilistic data association filter (JPDAF). The *inherited parallelism* of this algorithm enables it to be suitable for

implementation in a multiprocessor system. In this algorithm, the *a posteriori* probability of the origin of each measurement in the JPDAF is decomposed into two parts. The computation of one part becomes trivial and the algorithm developed here is implemented on the other part, which is shown to be *related to permanents*. The computational complexity of this algorithm has been analyzed in the worst case as well as in the average case. It has been concluded that this algorithm is more efficient than other existing ones in the average case. The results are fully documented in the very recent peer-reviewed publication [10].

Another important contribution emerging from this research is the development of a unified framework to provide a comprehensive understanding of the problem of data association in multitarget tracking. Specifically, the DFS algorithm in [9], which was developed under the sponsorship of the grant preceding this research, has been adapted for efficiently generating the data association hypotheses in the measurement-oriented and track-oriented approaches, where the total number of data association hypotheses is expected to increase drastically over that in the target-oriented approach. However, reduction in the overall computational cost may be feasible from observations on the conditional likelihood of data association hypotheses. In the target-oriented approach, the conditional likelihood of each data association hypothesis is unique. When targets are grouped into clusters, this uniqueness property does not hold for the measurement-oriented and track-oriented approaches. Two specialized DFS algorithms which can efficiently identify the data association hypotheses with identical conditional likelihood in the measurement-oriented and track-oriented approaches have been developed [11].

The problems of direction-of-arrival (DOA) estimation and parameter estimation of sinusoids in noise have been tackled extensively by subspace-based methods during the last decade ever since the harmonic retrieval method was proposed by Pisarenko in 1973 followed by Schmidt's doctoral dissertation in 1979 which led to a formal presentation of subspace-based methods in the open-literature in 1985. A new approach to parameter estimation based on signal-parameter-selectivity of the signal subspace (PESS) was proposed recently [12],[13]. This method exploits the structure of the steering matrix more thoroughly. The residual error is theoretically analyzed and it has been shown that in the PESS method this residual error is less than or equal to that of the LS-ESPRIT. In simulation, it is shown that the PESS method outperforms the TLS-ESPRIT method [14].

In DOA estimation and the related problem of parameter estimation of sinusoids in noise, the need for constraining the coefficients of a polynomial so that its roots fall on the unit circle occurs. While simple symmetry conditions suffice in a lot of cases [15], the need for more powerful constraints arises in low SNR situations. Investigations into such constraints lead to a set of necessary conditions on the coefficients of a polynomial for all its roots to lie on a unit circle [16]. The powerful mathematical structure built around the theory of resultants used in multidimensional systems theory for a variety of purposes [17], [18] are again used to attain the desired objective here.

Finally, to link image processing to target tracking for better tracker performance, a

recursive procedure based on total least squares (TLS) theory to reconstruct a high resolution image from a sequence of low resolution noisy and blurred frames (incorporating the inherent uncertainty in displacement estimations of the frames with respect to a reference frame) was developed in [19] as part of research conducted under the sponsorship of the previous ONR Grant N00014-86-K-0542. During the final phase of this research, a theoretical analysis was conducted to evaluate the robustness of the TLS algorithm developed in [19]. It was shown that with certain assumptions on the noise, the image reconstructed using the TLS algorithm has minimum variance with respect to all unbiased estimates. Furthermore, the quality of the reconstructed image improved with increase in the number of undersampled frames. In the case of blurred frames, higher resolution images may be reconstructed using the TLS algorithm with post-deblurring [20].

(c) References

[1] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Orlando, FL: Academic Press, Inc., 1988.

[2] D. B. Reid, "An algorithm for tracking multiple targets," IEEE Trans. Automatic Control, vol. AC-24, pp. 843-854, December 1979.

[3] Y. Bar-Shalom, *Multitarget-Multisensors Tracking: Advanced Applications*. Norwood, MA: Artech House Inc., 1990.

[4] R. J. Fitzgerald, "Development of practical PDA logic for multitarget tracking by microprocessor," in *Proc. American Controls Conference*, (Seattle, WA), pp. 889-898, June 1986.

[5] D. Sengupta and R. A. Iltis, "Neural solution to the multiple target tracking data association problem," IEEE Trans. Aerospace and Electronic Systems, vol. AES-25, pp. 96-108, January 1989.

[6] B. Zhou and N. K. Bose, "A comprehensive analysis of "neural solution to the multitarget tracking data association problem" IEEE Trans. Aerospace and Electronic Systems, vol. AES-29, pp. 260-263, January 1993.

[7] V. Nagarajan, M. R. Chidambara, and R. N. Sharma, "Combinatorial problems in multitarget tracking -- a comprehensive solution," IEE Proc. Part F, Communications, Radar and Signal Processing, vol. 134, pp. 113-118, February 1987.

[8] J. L. Fisher and D. P. Casasent, "Fast JPDA multitarget tracking algorithm," Applied Optics, vol. 28, pp. 371-376, January 1989.

[9] B. Zhou and N. K. Bose, "Multitarget tracking in clutter: fast algorithms for data association," IEEE Trans. Aerospace and Electronic Systems, vol. AES-29, pp. 352-363,

April 1993.

[10] B. Zhou and N. K. Bose, "An efficient algorithm for data association in multitarget tracking," IEEE Transactions on Aerospace and Electronic Systems, vol. 31, no. 1, January 1995, pp. 458-468.

[11] B. Zhou and N. K. Bose, "A unified approach to data association in multitarget tracking," Automatica, vol. 30, no. 9, 1994, pp. 1469-1472.

[12] J. T. Chun and N. K. Bose, "A novel subspace-based approach to parameter estimation," Digital Signal Processing, vol. 4, no. 1, January 1994, pp. 40-48.

[13] J. T. Chun and N. K. Bose, "Estimation via signal-selectivity of signal subspace (PESS) and its application to 2-D wavenumber estimation," Digital Signal Processing, vol. 5, no. 1, January 1995, pp. 58-76.

[14] J. T. Chun and N. K. Bose, "Parameter estimation by the signal-parameter-selectivity of the signal-subspace (PESS) method," preprint.

[15] J. Li and P. Stoica, "Efficient parameter estimation of partially polarized electromagnetic waves," IEEE Transactions on Signal Processing, vol. 40, no. 11, November 1994, pp. 314-325.

[16] N. K. Bose, "Improved method for parameter estimation of complex sinusoids in noise," (first presented at the International Symposium on Circuits and Systems, May 1993, Chicago, Illinois; IEEE Transactions on Circuits and Systems, Part 2, vol. 42, scheduled to appear.

[17] N. K. Bose, *Applied Multidimensional Systems Theory*, Van Nostrand Reinhold, New York, 1982.

[18] N. K. Bose and S. Basu, "Test for polynomial zeros on a polydisc distinguished boundary," IEEE Transactions on Circuits and Systems, 25, September 1978, pp. 684-693.

[19] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive total least squares algorithm for image reconstruction from noisy undersampled frames," Multidimensional Systems and Signal Processing, vol. 4, no. 3, July 1993, pp. 253-268.

[20] N. K. Bose, H. C. Kim, and B. Zhou, "Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames," Proceedings of International Conference on Image Processing, Austin, Texas, IEEE Computer Society Press, vol. III, November 1994, pp. 571-575.

3. Research Personnel

(a) Dr. N. K. Bose, HRB-Systems Professor of Electrical Engineering and Director of The Spatial and Temporal Signal Processing Center at The Pennsylvania State University, University Park, PA 16802.

Dr. Bose supervised the complete project, whose original funding expired on June 30, 1994. A no-cost extension permitted continuation of the grant to December 31, 1994. During the period of the grant, Dr. Bose supervised all phases of the research, provided new ideas for implementing data association efficiently, and prepared manuscripts for submission to and subsequent publication in reviewed journals. Dr. Bose also interacted with Professor Leon H. Sibul, of the Applied Research Laboratory at The Pennsylvania State University, and other scientists working on multitarget tracking, especially in the SDIO/IST Program. He regularly reported the progress of research at the annual SDIO/IST Workshop at Arlington, Virginia, where other principal investigators also came to present their ideas and share research results. He generated the leadership to link the focus of the research on design of fast algorithms for data association to peripheral tasks in tracking concerned with image processing and direction-of-arrival estimation from data collected at sensors.

(b) Bin Zhou, post-doctoral scholar at The Center for Multivariate Analysis and The Spatial and Temporal Signal Processing Center at The Pennsylvania State University, University Park, PA 16802.

Was supported as a post-doctoral scholar to work with the principal investigator on the design of fast algorithms for data association with parallel implementational capabilities. These algorithms were shown to perform more efficiently than existing ones on the average. He also helped in the formulation of a unified framework to tackle the multitarget tracking problem based on target-oriented, measurement-oriented, as well as track-oriented approaches.

(c) H. C. Kim, graduate research assistant at The Spatial and Temporal Signal Processing Center at The Pennsylvania State University, University Park, PA 16802.

Was partially supported because of his earlier research on image reconstruction from a sequence of undersampled, noisy, and blurred frames. This work can improve tracking performance by preprocessing collected data. With Dr. Bose and Dr. Zhou he developed theoretical tools for analyzing the performance of the total least-squares theory based recursive image reconstruction algorithm.

(d) J. Chun, graduate research assistant at The Spatial and Temporal Signal Processing Center at The Pennsylvania State University, University Park, PA 16802.

Was supported to work on the development of high performance algorithms for the direction-of-arrival estimation problem and the related problem of parameter estimation of sinusoids in noise. This research led to the development of a novel subspace-based method with proven advantages over existing methods, as mentioned elsewhere in this report and reported in detail in the referenced publications.

4. List of Publications

Peer-Reviewed Publications

1. N. K. Bose and L. H. Sibul, "Sensor array processing," The Electrical Engineering Handbook, ed. Richard C. Dorf, CRC Press, Inc., Boca Raton, Florida, 1993, pp. 359-367.

2. B. Zhou and N. K. Bose, "An efficient algorithm for data association in multitarget tracking," IEEE Transactions on Aerospace and Electronic Systems, vol. 31, no. 1, January 1995, pp. 458-468.

3. B. Zhou and N. K. Bose, "A unified approach to data association in multitarget tracking," Automatica, vol. 30, no. 9, 1994, pp. 1469-1472.

4. J. T. Chun and N. K. Bose, "A novel subspace-based approach to parameter estimation," Digital Signal Processing, vol. 4, no. 1, January 1994, pp. 40-48.

5. J. T. Chun and N. K. Bose, "Estimation via signal-selectivity of signal subspace (PESS) and its application to 2-D wavenumber estimation," Digital Signal Processing, vol. 5, no. 1, January 1995, pp. 58-76.

6. N. K. Bose, "Improved method for parameter estimation of complex sinusoids in noise," (first presented at the International Symposium on Circuits and Systems, May 1993, Chicago, IL); IEEE Transactions on Circuits and Systems, Part 2, vol. 42, scheduled to appear.

7. N. K. Bose, H. C. Kim, and B. Zhou, "Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames," Proceedings of International Conference on Image Processing, Austin, Texas, IEEE Computer Society Press, vol. III, pp. 571-575.

Presentations of Research Not Reported in Publications Above

8. N. K. Bose, "Algebraic methods for solving the sinusoidal modeling and parameter estimation problem in 1-D and 2-D," Seventh International Conference on Multivariate Analysis, Celebrating the Birth Centenary of P. C. Mahalanabis, New Delhi, India, December 1993, invited presentation.

9. J. Chun and N. K. Bose, "Estimation of 2-D wavenumbers," 37th IEEE Midwest Symposium on Circuits and Systems, Lafayette, Louisiana, August 3-5, 1994.

10. J. Chun and N. K. Bose, "Directions-of-arrival estimation for completely and partially correlated sources with radial basis function neural network," Conference on

Information Sciences and Systems, The Johns Hopkins University, Baltimore, Maryland, March 22-24, 1995.

Ph.D. Dissertations

A. H. C. Kim, "High Resolution Image Reconstruction from Undersampled Frames," Ph.D. Dissertation, Department of Electrical Engineering, The Pennsylvania State University, University Park, PA, May 1994.

B. J. Chun, "Subspace Methods for Direction-Finding in 3-D Space, Two-Dimensional Wavenumber Estimation, and Incomplete Image Restoration," Ph.D. Dissertation, Department of Electrical Engineering, The Pennsylvania State University, University Park, PA, scheduled for completion in May 1995.

5. Reports Distribution List

Scientific Officer Code: 1114SE
Rabinder N. Madan
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Administrative Grants Officer
Gerald Smith
Office of Naval Research Regional Director
Federal Building - Second Floor
Room 208
536 South Clark Street
Chicago, IL 60605-1588

Director, Naval Research Laboratory
Attn: Code 2657
Washington, DC 20375

Defense Technical Information Center
Building 5, Cameron Station
Alexandria, Virginia 22304-6145

SELECTED REPRINTS

References

R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Reading, Mass.: Addison-Wesley, 1987.

R. A. Haddad and T. W. Parsons, *Digital Signal Processing: Theory, Applications, and Hardware*, New York: Computer Science Press, 1991.

B. P. Horn, *Robot Vision*, Cambridge, Mass.: The MIT Press, 1986.

A. K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, N.J.: Prentice-Hall, 1989.

N. Jayant, "Signal compression: Technology targets and research directions," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 5, pp. 796–818, 1992.

H. G. Musmann, P. Pirsch, and H.-J. Grallert, "Advances in picture coding," *Proc. IEEE*, vol. 73, no. 4, pp. 523–548, 1985.

A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression*, New York: Plenum Press, 1988.

A. N. Netravali and J. D. Robbins, "Motion-compensated television coding: Part I," *Bell Syst. Tech. J.*, vol. 58, no. 3, pp. 631–670, 1979.

A. V. Oppenheim, A. S. Willsky, and I. T. Young, *Signals and Systems*, Englewood Cliffs, N.J.: Prentice-Hall, 1983.

W. F. Schreiber, *Fundamentals of Electronic Imaging Systems*, Berlin: Springer-Verlag, 1991.

Further Information

Other recommended sources of information include *IEEE Transactions on Circuits and Systems for Video Technology*, *IEEE Transactions on Image Processing*, and the *Proceedings of the IEEE*, April 1985, vol. 73.

16.3 Sensor Array Processing

N. K. Bose and L. H. Sibul

Multidimensional signal processing tools apply to aperture and sensor array processing. Planar sensor arrays can be considered to be sampled apertures. Three-dimensional or volumetric arrays can be viewed as multidimensional spatial filters. Therefore, the topics of sensor array processing, aperture processing, and multidimensional signal processing can be studied under a unified format. The basic function of the receiving array is transduction of propagating waves in the medium into electrical signals. Propagating waves are fundamental in radar, communication, optics, sonar, and geophysics. In electromagnetic applications, basic transducers are antennas and arrays of antennas. A large body of literature that exists on antennas and antenna arrays can be exploited in the areas of aperture and sensor array processing. Much of the antenna literature deals with transmitting antennas and their radiation patterns. Because of the reciprocity of transmitting and receiving transducers, key results that have been developed for transmitters can be used for analysis of receiver aperture and/or array processing. Transmitting transducers radiate energy in desired directions, whereas receiving apertures/arrays act as spatial filters that emphasize signals from a desired look direction while discriminating against interferences from other directions. The spatial filter **wavenumber** response is called the receiver beam pattern. Transmitting apertures are characterized by their radiation patterns.

Conventional beamforming deals with the design of fixed beam patterns for given specifications. Optimum beamforming is the design of beam patterns to meet a specified optimization criterion. It can be compared to optimum filtering, detection, and estimation. Adaptive **beamformers** sense their operating environment (for example, noise covariance matrix) and adjust beamformer parameters so that their performance is optimized [Monzingo and Miller, 1980]. Adaptive beamformers can be compared with adaptive filters.

Multidimensional signal processing techniques have found wide application in seismology—where a group of identical seismometers, called seismic arrays, are used for event location, studies

of the earth's sedimentation structure, and separation of coherent signals from noise, which sometimes may also propagate coherently across the array but with different horizontal velocities—by employing **velocity filtering** [Claerbout, 1976]. Velocity filtering is performed by multidimensional filters and allows also for the enhancement of signals which may occupy the same wavenumber range as noise or undesired signals do. In a broader context, beamforming can be used to separate signals received by sensor arrays based on frequency, wavenumber, and velocity (speed as well as direction) of propagation. Both the transfer and unit impulse-response functions of a velocity filter are two-dimensional functions in the case of one-dimensional arrays. The transfer function involves frequency and wavenumber (due to spatial sampling by equally spaced sensors) as independent variables, whereas the unit impulse response depends upon time and location within the array. Two-dimensional filtering is not limited to velocity filtering by means of seismic array. Two-dimensional spatial filters are frequently used, for example, in the interpretation of gravity and magnetic maps to differentiate between regional and local features. Input data for these filters may be observations in the survey of an area conducted over a planar grid over the earth's surface. Two-dimensional wavenumber digital filtering principles are useful for this purpose. Velocity filtering by means of two-dimensional arrays may be accomplished by properly shaping a three-dimensional response function $H(k_1,k_2,\omega)$. Velocity filtering by three-dimensional arrays may be accomplished through a four-dimensional function $H(k_1,k_2,k_3,\omega)$ as explained in the following subsection.

Spatial Arrays, Beamformers, and FIR Filters

A propagating plane wave, $s(\mathbf{x},t)$, is, in general, a function of the three-dimensional space variables and the time variable $(x_1,x_2,x_3) \triangleq \mathbf{x}$ and the time variable t. The 4-D Fourier transform of the stationary signal $s(\mathbf{x},t)$ is

$$S(\mathbf{k},\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(\mathbf{x},t)e^{-j(\omega t-\sum_{i=1}^{3}k_i x_i)}dx_1dx_2dx_3dt \tag{16.3}$$

which is referred to as the wavenumber–frequency spectrum of $s(\mathbf{x},t)$, and $(k_1,k_2,k_3) \triangleq \mathbf{k}$ denotes the wavenumber variables in radians per unit distance and ω is the frequency variable in radians per second. If c denotes the velocity of propagation of the plane wave, the following constraint must be satisfied

$$k_1^2 + k_2^2 + k_3^2 = \frac{\omega^2}{c^2}$$

If the 4-D Fourier transform of the unit impulse response $h(\mathbf{x},t)$ of a 4-D linear shift-invariant (LSI) filter is denoted by $H(k,\omega)$, then the response $y(\mathbf{x},t)$ of the filter to $s(\mathbf{x},t)$ is the 4-D linear convolution of $h(\mathbf{x},t)$ and $s(\mathbf{x},t)$, which is, uniquely, characterized by its 4-D Fourier transform

$$Y(\mathbf{k},\omega) = H(\mathbf{k},\omega)S(\mathbf{k},\omega) \tag{16.4}$$

The inverse 4-D Fourier transform, which forms a 4-D Fourier transform pair with Eq. (16.3), is

$$s(\mathbf{x},t) = \frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(\mathbf{k},\omega)e^{j(\omega t-\sum_{i=1}^{3}k_i x_i)}dk_1dk_2dk_3d\omega \tag{16.5}$$

It is noted that $S(\mathbf{k},\omega)$ in Eq. (16.3) is product separable, i.e., expressible in the form

$$S(\mathbf{k},\omega) = S_1(k_1)S_2(k_2)S_3(k_3)S_4(\omega) \tag{16.6}$$

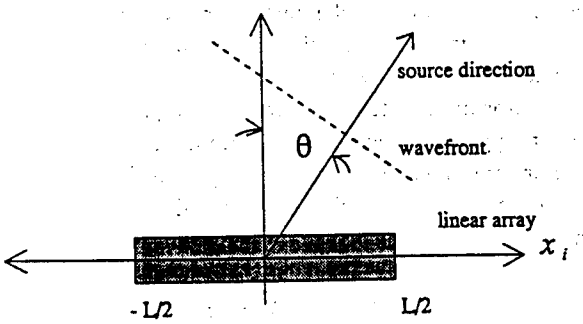


FIGURE 16.23 Uniformly weighted linear array.

where each function on the right-hand side is a univariate function of the respective independent variable, if and only if $s(\mathbf{x},t)$ in Eq. (16.3) is also product separable. In beamforming, $S_i(k_i)$ in Eq. (16.6) would be the far-field beam pattern of a linear array along the x_i-axis. For example, the normalized beam pattern of a uniformly weighted (shaded) linear array of length L is

$$S(k,\theta) = \frac{\sin\left(\dfrac{kL\sin\theta}{2}\right)}{\left(\dfrac{kL}{2}\sin\theta\right)} \tag{16.7}$$

where $\lambda = (2\pi/k)$ is the wavelength of the propagating plane wave and θ is the angle of arrival at array site as shown in Fig. 16.23. Note that θ is explicitly admitted as a variable in $S(k,\theta)$ to allow for the possibility that for a fixed wavenumber, the beam pattern could be plotted as a function of the angle of arrival. In that case, when θ is zero, the wave impinges the array broadside and the normalized beam pattern evaluates to unity.

The counterpart, in aperture and sensor array processing, of the use of window functions in spectral analysis for reduction of sidelobes is the use of aperture shading. In aperture shading, one simply multiplies a uniformly weighted aperture by the shading function. The resulting beam pattern is, then, simply the convolution of the beam pattern of the uniformly shaded volumetric array and the beam pattern of the shading function. Fourier transform relationship between the stationary signal $s(\mathbf{x},t)$ and the wavenumber frequency spectrum $S(\mathbf{k},\omega)$ allows one to exploit high-resolution spectral analysis techniques for the high-resolution estimation of the direction of arrival [Pillai, 1989]. The superscript *, t, and H denote, respectively, complex conjugate, transpose, and conjugate transpose.

Discrete Arrays for Beamforming

An array of sensors could be distributed at distinct points in space in various ways. Line arrays, planar arrays, and volumetric arrays could be either uniformly spaced or nonuniformly spaced, including the possibility of placing sensors randomly according to some probability distribution function. Uniform spacing along each coordinate axis permits one to exploit the well-developed multidimensional signal processing techniques concerned with filter design, DFT computation via FFT, and high-resolution spectral analysis of sampled signals [Dudgeon, 1977]. Nonuniform spacing sometimes might be useful for reducing the number of sensors, which otherwise might be constrained to satisfy a maximum spacing between uniformly placed sensors to avoid grating lobes due to aliasing, as explained later. A discrete array, uniformly spaced, is convenient for the synthe-

sis of a digital filter or beamformer by the performing of digital signal processing operations (namely delay, sum, and multiplication or weighting) on the signal received by a collection of sensors distributed in space. The sequence of the nature of operations dictates the types of beamformer. Common beamforming systems are of the straight summation, delay-and-sum, and weighted delay-and-sum types. The geometrical distribution of sensors and the weights w_i associated with each sensor are crucial factors in the shaping of the filter characteristics. In the case of a linear array of N equispaced sensors, which are spaced D units apart, starting at the origin $x_1 = 0$, the function

$$W(k_1) = \frac{1}{N}\sum_{n=0}^{N-1} w_n e^{-jk_1 nD} \tag{16.8}$$

becomes the **array pattern**, which may be viewed as the frequency response function for a finite impulse response (FIR) filter, characterized by the unit impulse response sequence $\{w_n\}$. In the case when $w_n = 1$, Eq. (16.8) simplifies to

$$W(k_1) = \frac{1}{N}\frac{\sin\left(\frac{k_1 ND}{2}\right)}{\sin\left(\frac{k_1 D}{2}\right)} \exp\left\{-j\frac{(N-1)k_1 D}{2}\right\} \tag{16.9}$$

If the N sensors are symmetrically placed on both sides of the origin, including one at the origin, and the sensor weights are $w_n = 1$, then the linear array pattern becomes

$$W(k_1) = \frac{1}{N}\frac{\sin\left(\frac{k_1 ND}{2}\right)}{\sin\left(\frac{k_1 D}{2}\right)}$$

For planar arrays, direct generalizations of the preceding linear array results can be obtained. To wit, if the sensors with unity weights are located at coordinates (kD, lD), where $k = 0, \pm 1, \pm 2, \ldots, \pm[(N-1)/2]$, and $l = 0, \pm 1, \pm 2, \ldots, \pm[(M-1)/2]$, for odd integer values of N and M, then the array pattern function becomes

$$W(k_1, k_2) = \frac{1}{NM}\sum_{k=-\left(\frac{N-1}{2}\right)}^{\left(\frac{N-1}{2}\right)}\sum_{l=-\left(\frac{M-1}{2}\right)}^{\left(\frac{M-1}{2}\right)} \exp\{-j(k_1 kD + k_2 lD)\}$$

$$= \frac{1}{NM}\frac{\sin\left(\frac{k_1 ND}{2}\right)}{\sin\left(\frac{k_1 D}{2}\right)}\frac{\sin\left(\frac{k_2 MD}{2}\right)}{\sin\left(\frac{k_2 D}{2}\right)} \tag{16.10}$$

Routine generalizations to 3-D spatial arrays are also possible. The array pattern functions for other geometrical distributions may also be routinely generated. For example, if unit weight sensors are located at the six vertices and the center of a regular hexagon, each of whose sides is D units long, then the array pattern function can be shown to be

$$W(k_1, k_2) = \frac{1}{7}\left[1 + 2\cos k_1 D + 4\cos\frac{k_1 D}{2}\cos\frac{\sqrt{3}k_2 D}{2}\right] \tag{16.11}$$

The array pattern function reveals how selective a particular beamforming system is. In the case of a typical array function shown in Eq. (16.9), the beamwidth, which is the width of the main lobe of the array pattern, is inversely proportional to the array aperture. Because of the periodicity of the array pattern function, the main lobe is repeated at intervals of $2\pi/D$. These repetitive lobes are called grating lobes, whose existence may be interpreted in terms of spatial frequency aliasing resulting from a sampling interval D due to the N receiving sensors located at discrete points in space. If the spacing D between sensors satisfies

$$D \le \frac{\lambda}{2} \tag{16.12}$$

where λ is the smallest wavelength component in the signal received by the array of sensors, then the grating lobes have no effect on the received signal. A plane wave of unit amplitude which is incident upon the array at bearing angle θ degrees, as shown in Fig. 16.23, produces outputs at the sensors given by the vector

$$\mathbf{s}(\theta) \triangleq \mathbf{s}_\theta = [\exp(j0)\ \exp(jk_1 D\sin\theta) \ldots \exp(jk_1(N-1)D\sin\theta)]^t \tag{16.13}$$

where $k_1 = 2\pi/\lambda$ is the wavenumber. In array processing, the array output y_θ may be viewed as the inner product of an array weight vector $\mathbf{w}$ and the steering vector $\mathbf{s}_\theta$. Thus, the beamformer response along a direction characterized by the angle θ is, treating $\mathbf{w}$ as complex,

$$y_\theta = \left\langle \mathbf{w}(\theta), \mathbf{s}_\theta \right\rangle = \sum_{k=0}^{N-1} w_k^* \exp(jk_1 kD\sin\theta) \tag{16.14}$$

The beamforming system is said to be robust if it performs satisfactorily despite certain perturbations [Ahmed and Evans, 1982]. It is possible for each component $s_{k\theta}$ of $\mathbf{s}_\theta$ to belong to an interval $[s_{k\theta} - \phi_{k\theta}, s_{k\theta} + \phi_{k\theta}]$, and a robust beamformer will require the existence of at least one weight vector $\mathbf{w}$ which will guarantee the output y_θ to belong to an output envelope for each $\mathbf{s}_\theta$ in the input envelope. The robust beamforming problem can be translated into an optimization problem, which may be tackled by minimizing the value of the array output power

$$P(\theta) = \mathbf{w}^H(\theta) R \mathbf{w}(\theta) \tag{16.15}$$

when the response to a unit amplitude plane wave incident at the steering direction θ is constrained to be unity, i.e., $\mathbf{w}^H(\theta)\mathbf{s}(\theta) = 1$, and R is the additive noise-corrupted signal autocorrelation matrix. The solution is called the minimum variance beamformer and is given by

$$\mathbf{w}_{MV}(\theta) = \frac{R^{-1}\mathbf{s}(\theta)}{\mathbf{s}^H(\theta)R^{-1}\mathbf{s}(\theta)} \tag{16.16}$$

and the corresponding power output is

$$P_{MV}(\theta) = \frac{1}{\mathbf{s}^H(\theta)R^{-1}\mathbf{s}(\theta)} \tag{16.17}$$

The minimum variance power as a function of θ can be used as a form of the data-adaptive estimate of the directional power spectrum. However, in this mode of solution, the coefficient vector is unconstrained except at the steering direction. Consequently, a signal tends to be regarded as an unwanted interference and is, therefore, suppressed in the beamformed output unless it is almost exactly aligned with the steering direction. Therefore, it is desirable to broaden the signal acceptance angle while at the same time preserving the optimum beamformer's ability to reject noise and interference outside this region of angles. One way of achieving this is by the application of the principle of superdirectivity.

Discrete Arrays and Polynomials

It is common practice to relate discrete arrays to polynomials for array synthesis purposes [Steinberg, 1976]. For volumetric equispaced arrays (it is only necessary that the spacing be uniform along each coordinate axis so that the spatial sampling periods D_i and D_j along, respectively, the ith and jth coordinate axes could be different for $i \neq j$), the weight associated with sensors located at coordinate (i_1D_1, i_2D_2, i_3D_3) is denoted by $w(i_1,i_2,i_3)$. The function in the complex variables $(z_1, z_2$ and $z_3)$ that is associated with the sequence $\{w(i_1,i_2,i_3)\}$ is the generating function for the sequence and is denoted by

$$W(z_1,z_2,z_3) = \sum_{i_1}\sum_{i_2}\sum_{i_3} w(i_1,i_2,i_3) z_1^{i_1} z_2^{i_2} z_3^{i_3} \tag{16.18}$$

In the electrical engineering and geophysics literature, the generating function $W(z_1,z_2,z_3)$ is sometimes called the z-transform of the sequence $\{w(i_1,i_2,i_3)\}$. When there are a finite number of sensors, a realistic assumption for any physical discrete array, $W(z_1,z_2,z_3)$ becomes a trivariate polynomial. In the special case when $w(i_1,i_2,i_3)$ is product separable, the polynomial $W(z_1,z_2,z_3)$ is also product separable. Particularly, this separability property holds when the shading is uniform, i.e., $w(i_1,i_2,i_3) = 1$. When the support of the uniform shading function is defined by $i_1 = 0, 1, \ldots, N_1 - 1$, $i_2 = 0, 1, \ldots, N_2 - 1$, and $i_3 = 0, 1, \ldots, N_3 - 1$, the associated polynomial becomes

$$W(z_1, z_2, z_3) = \sum_{i_1=0}^{N_1-1}\sum_{i_2=0}^{N_2-1}\sum_{i_3=0}^{N_3-1} z_1^{i_1} z_2^{i_2} z_3^{i_3} = \prod_{i=1}^{3} \frac{z_i^{N_i} - 1}{z_i - 1} \tag{16.19}$$

In this case, all results developed for the synthesis of linear arrays become directly applicable to the synthesis of volumetric arrays. For a linear uniform discrete array composed of N sensors with intersensor spacing D_1 starting at the origin and receiving a signal at a known fixed wavenumber k_1 at a receiving angle θ, the far-field beam pattern

$$S(k_1, \theta) \triangleq S(\theta) = \sum_{r=0}^{N-1} e^{jk_1 r D_1 \sin\theta}$$

may be associated with a polynomial $\sum_{r=0}^{N-1} z_1^r$, by setting $z_1 = e^{jk_1D_1\sin\theta}$. This polynomial has all its zeros on the unit circle in the z_1-plane. If the array just considered is not uniform but has a weighting factor w_r, for $r = 0, 1, \ldots, N_1 - 1$, the space factor,

$$Q(\theta) \triangleq \sum_{r=0}^{N_1-1} w_r e^{jk_1 D_1 r \sin\theta}$$

may again be associated with a polynomial $\sum_{r=0}^{N_1-1} w_r z_1^r$. By the pattern multiplication theorem, it is possible to get the polynomial associated with the total beam pattern of an array with weighted sensors by multiplying the polynomials associated with the array element pattern and the polynomial associated with the space factor $Q(\theta)$. The array factor $|Q(\theta)|^2$ may also be associated with the polynomial spectral factor

$$\left|Q(\theta)\right|^2 \leftrightarrow \sum_{r=0}^{N_1-1} w_r z_1^r \sum_{r=0}^{N_1-1} w_r^* (z_1^*)^r \tag{16.20}$$

where the weighting (shading) factor is allowed to be complex. Uniformly distributed apertures and uniformly spaced volumetric arrays which admit product separable sensor weightings can be treated by using the well-developed theory of linear discrete arrays and their associated polynomial. When the product separability property does not hold, scopes exist for applying results from multidimensional systems theory [Bose, 1982] concerning multivariate polynomials to the synthesis problem of volumetric arrays.

Velocity Filtering

Combination of individual sensor outputs in a more sophisticated way than the delay-and-sum technique leads to the design of multichannel velocity filters for linear and planar as well as spatial arrays. Consider, first, a linear (1-D) array of sensors, which will be used to implement velocity discrimination. The pass and rejection zones are defined by straight lines in the (k_1,ω)-plane, where

$$k_1 = \frac{\omega}{V} = \frac{\omega}{(v/\sin\theta)}$$

is the wavenumber, ω the angular frequency in radians/second, V the apparent velocity on the earth's surface along the array line, v the velocity of wave propagation, and θ the horizontal arrival direction. The transfer function

$$H(\omega,k_1) = \begin{cases} 1, & -\dfrac{|\omega|}{V} \le k_1 \le \dfrac{|\omega|}{V} \\ 0, & \text{otherwise} \end{cases}$$

of a "pie-slice" or "fan" velocity filter [Bose, 1985] rejects totally wavenumbers outside the range $-|\omega|/V \le k_1 \le |\omega|/V$ and passes completely wavenumbers defined within that range. Thus, the transfer function defines a high-pass filter which passes signals with apparent velocities of magnitude greater than V at a fixed frequency ω. If the equispaced sensors are D units apart, the spatial sampling results in a periodic wavenumber response with period $k_1 = 1/(2D)$. Therefore, for a specified apparent velocity V, the resolvable wavenumber and frequency bands are, respectively, $-1/(2D) \le k_1 \le 1/(2D)$ and $-V/(2D) \le \omega \le V/(2D)$ where $\omega/(2D)$ represents the folding frequency in radians/second.

Linear arrays are subject to the limitation that the source is required to be located on the extended line of sensors so that plane wavefronts approaching the array site at a particular velocity excite the individual sensors, assumed equispaced, at arrival times which are also equispaced. In seismology, the equispaced interval between successive sensor arrival times is called a move-out or step-out and equals $(D \sin\theta)/v = D/V$. However, when the sensor-to-source azimuth varies, two or more independent signal move-outs may be present. Planar (2-D) arrays are then required to discriminate between velocities as well as azimuth. Spatial (3-D) arrays provide additional scope to

the enhancement of discriminating capabilities when sensor/source locations are arbitrary. In such cases, an array origin is chosen and the mth sensor location is denoted by a vector $(x_{1m}\, x_{2m}\, x_{3m})^t$ and the frequency wavenumber response of an array of sensors is given by

$$H(\omega, k_1, k_2, k_3) = \frac{1}{N}\sum_{m=1}^{N} H_m(\omega)\exp\left[\sum_{i=1}^{3} -j2\pi k_i x_{im}\right]$$

where $H_m(\omega)$ denotes the frequency response of a filter associated with the mth recording device (sensor). The sum of all N filters provides flat frequency response so that waveforms arriving from the estimated directions of arrival at estimated velocities are passed undistorted and other waveforms are suppressed. In the planar specialization, the 2-D array of sensors leads to the theory of 3-D filtering involving a transfer function in the frequency wavenumber variables f, k_1, and k_2. The basic design equations for the optimum, in the least-mean-square error sense, frequency wavenumber filters have been developed [Burg, 1964]. This procedure of Burg can be routinely generalized to the 4-D filtering problem mentioned above.

Acknowledgment

N. K. Bose and L. H. Sibul acknowledge the support provided by the Office of Naval Research under, respectively, Contract N00014-92-J-1755 and the Fundamental Research Initiatives Program.

Defining Terms

Array pattern: Fourier transform of the receiver weighting function taking into account the positions of the receivers.

Beamformers: Systems commonly used for detecting and isolating signals that are propagating in a particular direction.

Grating lobes: Repeated main lobes in the array pattern interpretable in terms of spatial frequency aliasing.

Velocity filtering: Means for discriminating signals from noise or other undesired signals because of their different apparent velocities.

Wavenumber: 2π (spatial frequency in cycles per unit distance).

References

K. M. Ahmed and R. J. Evans, "Robust signal and array processing," *IEE Proceedings, F: Communications, Radar, and Signal Processing*, vol. 129, no. 4, pp. 297–302, 1982.

N. K. Bose, *Applied Multidimensional Systems Theory*, New York: Van Nostrand Reinhold, 1982.

N. K. Bose, *Digital Filters*, New York: Elsevier Science North-Holland, 1985. Reprint ed., Malabar, Fla.: Krieger Publishing, 1993.

J. P. Burg, "Three-dimensional filtering with an array of seismometers," *Geophysics*, vol. 23, no. 5, pp. 693–713, 1964.

J. F. Claerbout, *Fundamentals of Geophysical Data Processing*, New York: McGraw-Hill, 1976.

D. E. Dudgeon, "Fundamentals of digital array processing," *Proc. IEEE*, vol. 65, pp. 898–904, 1977.

R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, New York: Wiley, 1980.

S. M. Pillai, *Array Signal Processing*, New York: Springer-Verlag, 1989.

B. D. Steinberg, *Principles of Aperture and Array System Design*, New York: Wiley, 1976.

An Efficient Algorithm for Data Association in Multitarget Tracking

B. ZHOU, Member, IEEE

N. K. BOSE, Fellow, IEEE
The Pennsylvania State University

An efficient algorithm is developed to compute the *a posteriori* probabilities of the origins of measurements in the joint probabilistic data association filter (JPDAF). The inherited parallelism of this algorithm enables it to be suitable for implementation in a multiprocessor system. In this algorithm, the *a posteriori* probability of the origin of each measurement in the JPDAF is decomposed into two parts. The computation of one part becomes trivial and the algorithm developed here is implemented on the other part, which is shown to be related to permanents. The computational complexity of this algorithm is analyzed in the worst case as well as in the average case. An analysis of this algorithm enables us to conclude that this algorithm is more efficient than other existing ones in the average case.

Manuscript received September 1, 1992; revised December 10, 1993.

IEEE Log No. T-AES/31/1/08014.

This work was supported by a grant from SDIO/IST and managed by the Office of Naval Research under Contracts N00014-86-0542 and N00014-92-J-1755.

Authors' current addresess: B. Zhou, 102 Malott Hall, Cornell University, Ithaca, NY 14853; N. K. Bose, The Spatial and Temporal Signal Processing Center, Department of Electrical and Computer Engineering, Pennsylvania State University, University Park, PA 16802.

I. INTRODUCTION

In a cluttered environment, the received measurements may not all arise from the targets of interest. Some of them may be from clutter or false alarm. As a result, there always exist ambiguities in the association between the previous known targets and measurements. To solve the problem of data association between targets and measurements, three typical approaches have been reported in the literature. The first is called a *target-oriented* approach in which each measurement is assumed to have originated from either a known target or clutter, as in the *joint probabilistic data association filter* (JPDAF) [1, 2]. The second is called a *measurement-oriented* approach in which each measurement is hypothesized to have originated from either a known target, a new target, or clutter [3]. The third approach is referred to as *track-oriented* where each track is hypothesized to be either undetected, terminated, associated with a measurement, or linked to the start of a maneuver [4–6]. In these approaches, the number of data association hypotheses could increase rapidly with the increase in the number of targets and the number of measurements. Therefore, in a multitarget tracking algorithm, the computational cost in generating the data association hypotheses would be excessive when the number of targets and the number of measurements are large.

In recent years, a lot of attention has been given to the task of improving the computational efficiency of a multitarget tracking algorithm. Fitzgerald [7] proposed an *ad hoc* formula to approximately compute the β_j^ts in the JPDAF. For $j > 0$, β_j^t is the *a posteriori* probability that measurement j originated from target t. For $j = 0$, β_0^t is the *a posteriori* probability that no measurement originated from target t. Fitzgerald's formula breaks down when four targets are required to be tracked [8]. Sengupta and Iltis [8, 9] developed an analog neural network to emulate the JPDAF. They showed that the neural network is capable of handling two to six targets and three to twenty measurements. However, the heuristic nature of their approach makes implementation difficult [10]. Alternatively, Nagarajan, et al [11], arranged the hypotheses in the measurement-oriented approach [3] in a special order so that the probabilities of the hypotheses are proportional to the product of certain probability factors already evaluated. The algorithm locates the N globally best hypotheses without evaluating all of them. In [12], Fisher and Casasent developed a fast joint probabilistic data association (JPDA) algorithm. In their algorithm, the computation of the β_j^ts was implemented using enormous number of vector inner product operations. Since the vector inner product operation could be easily realized on an optical processor, they proposed a specialized optical processor to approximately implement the

fast JPDA algorithm. Recently, Zhou and Bose [13] proposed a depth-first search (DFS) approach to efficiently compute the β_j^ts in the JPDAF. Their algorithm requires much less computation than the fast JPDA algorithm in the average case, when there are, on average, two to three measurements inside the validation gate of a target. The validation gate of a target is a region with a prescribed size around the predicted position of the target.

As shown in [13, 14], the performance of an approximation of the JPDAF degrades drastically when the density of targets is high. Therefore, it is important to implement the JPDAF accurately in a dense target environment. Among the algorithms mentioned above, only the two algorithms proposed in [12, 13] have the potential for accurately computing the β_j^ts in the JPDAF. The fast JPDA algorithm in [12] is suitable for implementation in a multiprocessor system. Nevertheless, the computational cost is very high when the number of targets and the number of measurements are large since the design of the algorithm is based on the worst case scenario. In the worst case, all measurements fall inside the intersection of the validation gates of all targets. Although the DFS algorithm [13] is much more efficient than the fast JPDA algorithm in the average case, it is specifically directed towards implementation through a system with a powerful centralized processor, normally used in ground-based tracking systems. Our objective here is to propose an algorithm which not only performs better than the DFS algorithm in the average case, but which is also suitable for implementation in a tracking system with several spatially distributed microprocessors.

A brief review of the JPDAF is given in Section II, followed by the problem formulation in Section III. The new algorithm is developed in Section IV. In Section V, the computational complexity of the new algorithm is analyzed in the worst case as well as in the average case. Finally, the advantages of the new algorithm are summarized in Section VI.

II. REVIEW OF THE JPDAF

We assume that there are n targets being tracked at time index k. Let the dynamic model for target t be described by

$$x^t(k+1) = F^t(k)x^t(k) + G^t(k)w^t(k) \quad (1)$$

$$z^t(k) = H^t(k)x^t(k) + v^t(k), \qquad t = 1,2,\ldots,n \quad (2)$$

where $x^t(k)$ is the N^t-dimensional state vector, $z^t(k)$ is the M^t-dimensional measurement vector with M^t actually independent of t, $F^t(k)$, $G^t(k)$ and $H^t(k)$ are known model matrices, $w^t(k)$ and $v^t(k)$ are, respectively, the p^t and M^t-dimensional noise vectors, which are assumed to be zero-mean independent identically distributed Gaussian processes with known covariances:

$$E\{w^t(k)(w^t(j))'\} = Q^t(k)\delta_{kj} \quad (3)$$

$$E\{v^t(k)(v^t(j))'\} = R^t(k)\delta_{kj} \quad (4)$$

where $\delta_{kj} = 1$ if $k = j$ and $\delta_{kj} = 0$ otherwise. The prime superscript indicates matrix transposition. The initial target state $x^t(0)$ is assumed to be normally distributed with mean $\hat{x}^t(0 \mid 0)$ and covariance $P^t(0 \mid 0)$. It is also assumed to be independent of $w^t(k)$ and $v^t(k)$ for all $k \geq 0$.

Suppose that m measurements are received at time index k. In a cluttered environment, m does not necessarily equal n and it may be difficult to distinguish whether a measurement originated from a target or from clutter. It is reasonable to denote a measurement at time index k by

$$z(k) = \begin{cases} z^t(k) & \text{if } z(k) \text{ is from target } t \\ z^c(k) & \text{if } z(k) \text{ is from clutter.} \end{cases} \quad (5)$$

The measurement $z^c(k)$ is usually assumed to be uniformly distributed in the surveillance region and the number of clutter is subject either to Poisson distribution or to uniform distribution [1]. In this paper, Poisson distribution is assumed for the purpose.

In the JPDAF, the *a posteriori* probability β_j^t is computed. For $j > 0$, β_j^t is the *a posteriori* probability that validated measurement j originated from target t. β_0^t is the *a posteriori* probability that no measurement originated from target t. A validated measurement is one which is either inside or on the boundary of the validation gate of a target. Mathematically, a validation gate is defined by

$$(z(k) - \hat{z}^t(k))'S^t(k)^{-1}(z(k) - \hat{z}^t(k)) \leq g^2 \quad (6)$$

where $\hat{z}^t(k)$ is the predicted value of $z(k)$ for target t. The error, $(z(k) - \hat{z}^t(k))$, is the innovation generated from $z(k)$ for target t, $S^t(k)$ is its covariance matrix, and g is a selected threshold. As discussed in [2, 15], the choice $g > \sqrt{M^t} + 2$ ensures that the correct measurements will lie within the gate with probability 0.999. The dimension of $z(k)$ is M^t. The inequality given in (6) is said to generate a validation test. The result of the test is kept in what is called a validation matrix Ω. This validation matrix Ω is a $m \times (n+1)$ rectangular matrix defined as [2],

$$\Omega = [\omega_{jt}] = \begin{array}{c} \overbrace{}^{t} \\ \begin{matrix} 0 & 1 & 2 & \cdots & n \end{matrix} \\ \begin{pmatrix} \omega_{10} & \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{20} & \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \omega_{m0} & \omega_{m1} & \omega_{m2} & \cdots & \omega_{mn} \end{pmatrix} \begin{matrix} 1 \\ 2 \\ \vdots \\ m \end{matrix} \end{array} \} j \quad (7)$$

where $\omega_{j0} = 1$ for $j = 1,2,\ldots,m$, $\omega_{jt} = 1$ if measurement j is inside the validation gate of target t and $\omega_{jt} = 0$ otherwise for $j = 1,2,\ldots,m$, and $t = 1,2,\ldots,n$. Based on the validation matrix Ω, data association hypotheses (or feasible events $\mathcal{E}$s [2]) are generated subject to the following two restrictions:

1) Each measurement can have only one origin (either a specific target or clutter).

2) No more than one measurement originates from a target.

This leads to a combinatorial problem where the number of data association hypotheses increases exponentially with n and m.

Each feasible event $\mathcal{E}$ is represented by a hypothesis matrix $\hat{\Omega}$ in [2]. $\hat{\Omega}$ has the same size as the validation matrix Ω. A typical element in $\hat{\Omega}$ is denoted by $\hat{\omega}_{jt}$ where

$$\hat{\omega}_{jt} = \begin{cases} 1 & \text{if } \omega_{jt} = 1 \text{ and measurement } j \text{ is hypothesized to be from clutter for } t = 0 \\ 1 & \text{if } \omega_{jt} = 1 \text{ and measurement } j \text{ is hypothesized to be associated with target } t \text{ for } t \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Corresponding to the two restrictions above, in $\hat{\Omega}$, there is at most one unit element in each column except for $t = 0$ and there is exactly one unit element in each row.

After each $\hat{\Omega}$ is obtained, the conditional probability of the corresponding data association hypothesis or feasible event is calculated by a formula given in [2]. A simplified version of this formula is given as

$$P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z}) = \frac{1}{c}(P_0)^{\min(n,m)-m_d} \prod_{j:\hat{\omega}_{jt}=1} P_j^t \tag{9}$$

$$\text{for} \quad j = 1,2,\ldots,m, \quad \text{and} \quad t = 1,2,\ldots,n$$

where $\mathcal{Z}$ is the set of all measurements received up to current time index k, c is a normalizing constant, m_d is the number of targets detected in this feasible event $\mathcal{E}$, $\hat{\omega}_{jt} = 1$ indicates that measurement j is associated with target t in the event, and

$$P_j^t = \begin{cases} N(\tilde{z}_j^t; 0, S^t) P_D & \text{if } \omega_{jt} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$P_0^t = \lambda(1 - P_D) = P_0 \tag{11}$$

$$\text{for} \quad j = 1,2,\ldots,m, \quad \text{and} \quad t = 1,2,\ldots,n.$$

In (10) and (11), λ is the clutter density, P_D is the probability of detection, and $N(\tilde{z}_j^t; 0, S^t)$ is a normal density function having zero mean and a covariance matrix S^t with

$$\tilde{z}_j^t(k) = z_j(k) - H^t(k)\hat{x}^t(k \mid k-1). \tag{12}$$

In (12), $H^t(k)$ is the observation matrix for target t as defined in (2) and $\hat{x}^t(k \mid k-1)$ is the one-step prediction of the state estimate of target t. It is understood that the normalizing constant c in (9) is obtained by the summation, $\sum_{\mathcal{E}(\hat{\Omega})} P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z})$. Therefore, c is omitted hereafter. The *a posteriori* probability β_j^t is computed from the conditional probabilities in (9) by

$$\beta_j^t = \sum_{\mathcal{E}(\hat{\Omega})} P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z}) \hat{\omega}_{jt} \tag{13}$$

$$\beta_0^t = 1 - \sum_{j=1}^{m} \beta_j^t \tag{14}$$

$$\text{for} \quad j = 1,2,\ldots,m, \quad \text{and} \quad t = 1,2,\ldots,n.$$

The equations of the JPDAF are the same as those of a standard Kalman filter with the following two changes.

1) The innovation vector is generated by using relation:

$$\tilde{z}^t(k) = \sum_{j=1}^{m} \beta_j^t(k) \tilde{z}_j^t(k). \tag{15}$$

2) The update equation for the covariance matrix $P^t(k,k \mid k)$ is changed to

$$\begin{aligned} P^t(k,k \mid k) = {} & \beta_0^t(k) P^t(k,k \mid k-1) \\ & + (1 - \beta_0^t(k)) P^{t*}(k,k \mid k) \\ & + K^t(k) W^t(k) K^t(k)' \end{aligned} \tag{16}$$

where $K^t(k)$ is the Kalman gain for target t,

$$P^{t*}(k,k \mid k) = [I - K^t(k)H^t(k)]P^t(k,k \mid k-1) \tag{17}$$

denotes the covariance update for a single correct return, and

$$W^t(k) = \sum_{j=1}^{m} \beta_j^t(k) \tilde{z}_j^t(k) \tilde{z}_j^t(k)' - \tilde{z}^t(k) \tilde{z}^t(k)'. \tag{18}$$

In general, a covariance matrix $P^t(p,q \mid r)$ is defined as

$$P^t(p,q \mid r) = E\{[x^t(p) - \hat{x}^t(p \mid r)][x^t(q) - \hat{x}^t(q \mid r)]'\} \tag{19}$$

where $\hat{x}^t(p \mid r) = E\{x(p) \mid \text{measurements received up to time index } r\}$.

In the next section, the computation of the β_j^ts is reformulated in such a way that the construction of the hypothesis matrices $\hat{\Omega}$ is not necessary.

III. PROBLEM FORMULATION

In the original JPDAF [2], the computation of β_j^t consists of the following three steps.

1) Construct a validation matrix Ω and compute P_j^t using (10) and (11).

2) Generate data association hypotheses and compute $P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z})$ using (9).

3) Compute β_j^t using (13) and (14).

In this section, each β_j^t is decomposed into two parts so that its computations does not require the generation of the data association hypotheses or the construction of the hypothesis matrices as in the original JPDAF [2] and the DFS algorithms [13]. Let $\mathcal{A}$ denote an ordered set of nonzero integers associated with the targets. Let $\mathcal{B}$ denote another ordered set of non-zero integers which are the indices of measurements. The symbol 0 is the index for clutter. Suppose that there are n targets and m measurements are received at a certain time instant. Then,

$$\mathcal{A} = \{1,2,\ldots,n\} \tag{20}$$

$$\mathcal{B} = \{0,1,\ldots,m\}. \tag{21}$$

With the above notation, (13) may be rewritten as

$$\beta_j^t(\mathcal{A},\mathcal{B}) = \sum_{\mathcal{E}(\hat{\Omega})} P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z}) \hat{\omega}_{jt} \tag{22}$$

$$\text{for} \quad j = 1,2,\ldots,m, \quad \text{and} \quad t = 1,2,\ldots,n.$$

Since P_j^t is a factor of the $P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z})$s when $\hat{\omega}_{jt} = 1$, β_j^t may be decomposed as

$$\beta_j^t(\mathcal{A},\mathcal{B}) = P_j^t F(\mathcal{A}\backslash t, \mathcal{B}\backslash j) \tag{23}$$

$$\text{for} \quad j = 1,2,\ldots,m, \quad \text{and} \quad t = 1,2,\ldots,n$$

where the difference set $\mathcal{C}\backslash l$ is the set derived from $\mathcal{C}$ by deleting element l. In (23), $F(\mathcal{A}\backslash t, \mathcal{B}\backslash j)$ is a function of P_l^τs for $l \neq j$ and $\tau \neq t$. An interpretation of $F(\mathcal{A}\backslash t, \mathcal{B}\backslash j)$ may be obtained using the definition of $P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z})$ [2].

For $j = 1,2,\ldots,m$, and $t = 0,1,2,\ldots,n$, let $\mathcal{E}_{jt}$ denote the hypothesis that measurement j originated either from target t if $t \neq 0$ or from clutter if $t = 0$. Then, each feasible event $\mathcal{E}$ may be written as [2]

$$\mathcal{E}(\hat{\Omega}) = \bigcap_{j=1}^{m} \mathcal{E}_{jp(t)}(\hat{\omega}_{jp(t)}) \tag{24}$$

where $p(t)$ denotes an element of the set of permutations of $\{t\} \triangleq \{0,1,2,\ldots,n\}$. Furthermore, let $\hat{\Omega}'_{(jt)}$ be the hypothesis matrix obtained from $\hat{\Omega}$ after eliminating the jth row and the tth column. $P(\mathcal{E}_{jt}(\hat{\omega}_{jt}) \mid \mathcal{Z})$ is the probability of the hypothesis that measurement j originated from target t subject to the assumption that data association between the remaining measurements and remaining targets is ignored, and $\sum_{\mathcal{E}(\hat{\Omega}'_{(jt)})} P(\mathcal{E}(\hat{\Omega}'_{(jt)}) \mid \mathcal{E}_{jt}(\hat{\omega}_{jt}), \mathcal{Z})$ denotes the probability of all the data association hypotheses among the remaining measurements and targets conditioned on the hypothesis that measurement j is associated with target t. Since all the feasible events $\mathcal{E}$s are mutually exclusive, (22) can be rearranged as below.

$$\begin{aligned}
\beta_j^t(\mathcal{A},\mathcal{B}) &= \sum_{\mathcal{E}(\hat{\Omega})} P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z}) \hat{\omega}_{jt} \\
&= P\left(\bigcup_{\mathcal{E}(\hat{\Omega}):\hat{\omega}_{jt}=1} \mathcal{E}(\hat{\Omega}) \mid \mathcal{Z} \right) \\
&= P\left(\mathcal{E}_{jt}(\hat{\omega}_{jt}) \Leftrightarrow \left(\bigcup_{\mathcal{E}(\hat{\Omega}):\hat{\omega}_{jt}=1} \mathcal{E}(\hat{\Omega}) \right) \mid \mathcal{Z} \right) \\
&= P(\mathcal{E}_{jt}(\hat{\omega}_{jt}) \mid \mathcal{Z}) \\
&\quad \times P\left(\bigcup_{\mathcal{E}(\hat{\Omega})} \left(\bigcap_{\substack{l \neq j \\ \tau \neq t}} \mathcal{E}_{l\tau}(\hat{\omega}_{l\tau}) \right) \mid \mathcal{E}_{jt}(\hat{\omega}_{jt}), \mathcal{Z} \right) \\
&= P(\mathcal{E}_{jt}(\hat{\omega}_{jt}) \mid \mathcal{Z}) P\left(\bigcup_{\mathcal{E}(\hat{\Omega}'_{(jt)})} \mathcal{E}(\hat{\Omega}'_{(jt)}) \mid \mathcal{E}_{jt}(\hat{\omega}_{jt}), \mathcal{Z} \right) \\
&= P(\mathcal{E}_{jt}(\hat{\omega}_{jt}) \mid \mathcal{Z}) \sum_{\mathcal{E}(\hat{\Omega}'_{(jt)})} P(\mathcal{E}(\hat{\Omega}'_{(jt)}) \mid \mathcal{E}_{jt}(\hat{\omega}_{jt}), \mathcal{Z})
\end{aligned}$$

$$\text{for} \quad j = 1,2,\ldots,m, \quad \text{and} \quad t = 1,2,\ldots,n. \tag{25}$$

Comparing (25) with (23), we have

$$P_j^t = P(\mathcal{E}_{jt}(\hat{\omega}_{jt}) \mid \mathcal{Z}) \tag{26}$$

$$F(\mathcal{A}\backslash t, \mathcal{B}\backslash j) = \sum_{\mathcal{E}(\hat{\Omega}'_{(jt)})} P(\mathcal{E}(\hat{\Omega}'_{(jt)}) \mid \mathcal{E}_{jt}(\hat{\omega}_{jt}), \mathcal{Z}). \tag{27}$$

Therefore, $F(\mathcal{A}\backslash t, \mathcal{B}\backslash j)$ in (23) contains the sum of the conditional probabilities of the data association hypotheses among all the targets, the measurements, and clutter, given that measurement j is associated with target t. In the case of the probabilistic data association filter (PDAF) [15], where P_j^t differs from the *a posteriori* probability, β_j^t, by a normalization constant, $F(\mathcal{A}\backslash t, \mathcal{B}\backslash j)$ may be considered to be due to the interference from other targets on β_j^t if (23) is compared with its counterpart in the PDAF.

Similarly, β_0^t may be decomposed as

$$\beta_0^t = P_0 F(\mathcal{A}\backslash t, \mathcal{B}), \qquad \text{for} \quad t = 1,2,\ldots,n \tag{28}$$

where $F(\mathcal{A}\backslash t, \mathcal{B})$ contains the sum of the conditional probabilities of the data association hypotheses among all the targets, the measurements, and clutter, given that target t is not associated with any measurement. In (28), $F(\mathcal{A}\backslash t, \mathcal{B})$ may also be considered as the interference on β_0^t from other targets. The computation

of β_j^t after reformulation can now be summarized in three steps as given below.

1) Construct a validation matrix Ω and compute P_j^t using (10) and (11).
2) Compute $F(\mathcal{A}\backslash t,\mathcal{B}\backslash j)$ and $F(\mathcal{A}\backslash t,\mathcal{B})$.
3) Compute β_j^t using (23) and (28).

To compute β_0^t for $t = 1,\ldots,n$, using (14) instead of (28), the β_j^ts need to be normalized. The β_j^ts computed using (23) and (28) are not normalized. The normalizing constant c is determined by the following relationship.

$$\sum_{j=0}^{m}\beta_j^t(\mathcal{A},\mathcal{B}) = c, \qquad \text{for} \quad t = 1,2,\ldots,n. \tag{29}$$

Suppose that c is computed in the $t = 1$ case as

$$\sum_{j=0}^{m}\beta_j^1(\mathcal{A},\mathcal{B}) = c. \tag{30}$$

The same constant also applies, in all related cases where $t \neq 1$. Then, β_j^1 for $j = 0,1,\ldots,m$, are normalized as below

$$\beta_j^1 \leftarrow \frac{1}{c}\beta_j^1. \tag{31}$$

Using c, β_j^t, for $t = 2,\ldots,n$, and $j = 1,\ldots,m$, can also be normalized as the way the β_j^1s do. Now, β_0^t for $t = 2,\ldots,n$, are ready to be computed using (14). In the next section, an algorithm is developed to compute $F(\mathcal{A}\backslash t,\mathcal{B}\backslash j)$ for $t = 1,\ldots,n$, and $j = 1,\ldots,m$, and $F(\mathcal{A}\backslash 1,\mathcal{B})$.

IV. ALGORITHM DEVELOPMENT

From the notation introduced in the last section, $F(\mathcal{A},\mathcal{B})$ denotes the sum of the conditional probabilities of the data association hypotheses among the targets in $\mathcal{A}$ and the measurements in $\mathcal{B}$. Using (22) and (27), we have

$$F(\mathcal{A},\mathcal{B}) = \sum_{\mathcal{E}(\hat{\Omega})} P(\mathcal{E}(\hat{\Omega}) \mid \mathcal{Z}) = \sum_{j=0}^{m}\beta_j^t(\mathcal{A},\mathcal{B}). \tag{32}$$

After substituting (23) and (28) into (32), a recurrence relation for $F(\mathcal{A},\mathcal{B})$ can be obtained. For $t \in \mathcal{A}$,

$$F(\mathcal{A},\mathcal{B}) = P_0F(\mathcal{A}\backslash t,\mathcal{B}) + \sum_{j\in\mathcal{B}\backslash 0} P_j^tF(\mathcal{A}\backslash t,\mathcal{B}\backslash j). \tag{33}$$

The algorithm for computing $F(\mathcal{A},\mathcal{B})$ is obtained by recursive implementation of (33). The initial conditions associated with (33) are given below.

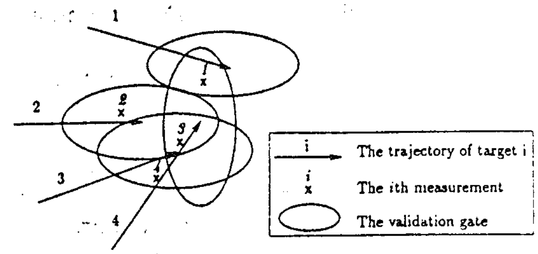


Fig. 1. Typical layout of 4 targets and 4 measurements.

1) If t is the only element in $\mathcal{A}$, then,

$$F(\mathcal{A},\mathcal{B}) = \sum_{j\in\mathcal{B}} P_j^t. \tag{34}$$

Note that $P_0^t = P_0$ in the above equation according to the notation introduced in (11).

2) If $j = 0$ is the only element in $\mathcal{B}$, then,

$$F(\mathcal{A},\mathcal{B}) = (P_0)^{|\mathcal{A}|} \tag{35}$$

where $|\mathcal{A}|$ denotes the number of elements in $\mathcal{A}$.

3) If $|\mathcal{B}| = 2$ and $|\mathcal{A}| \geq |\mathcal{B}|$, then,

$$F(\mathcal{A},\mathcal{B}) = (P_0)^{|\mathcal{A}|-1}\left(P_0 + \sum_{\substack{t\in\mathcal{A}\\ j\in\mathcal{B}\backslash 0}} P_j^t\right). \tag{36}$$

In order to demonstrate how the recursive algorithm works, consider an example.

EXAMPLE. Suppose that there are 4 closely spaced targets under surveillance. On one radar scan, four measurements are received as shown in Fig. 1. Then,

$$\mathcal{A} = \{1,2,3,4\} \tag{37}$$

$$\mathcal{B} = \{0,1,2,3,4\}. \tag{38}$$

The conventional validation matrix occurring in the JPDAF for this example is given in (39)

$$\Omega = \overbrace{\begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array}}^{t} \atop \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix} \left.\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}\right\} j. \tag{39}$$

For target 1, there is only one measurement which falls inside its validation gate. Therefore, only β_0^1 and β_1^1 are none zero. Using (23) and (28), β_0^1 and β_1^1 may be computed as

$$\beta_0^1(\mathcal{A},\mathcal{B}) = P_0F(\mathcal{A}\backslash 1,\mathcal{B}) \tag{40}$$

$$\beta_1^1(\mathcal{A},\mathcal{B}) = P_1^1F(\mathcal{A}\backslash 1,\mathcal{B}\backslash 1) \tag{41}$$

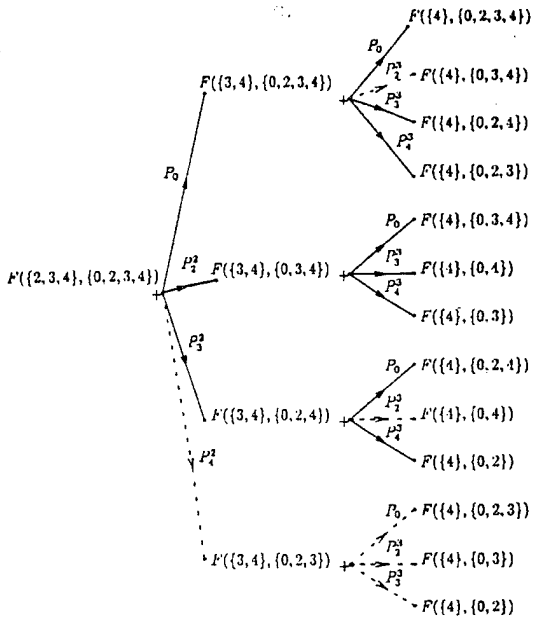


Fig. 2. Illustration of example using recursive algorithm.

where

$$\mathcal{A}\backslash 1 = \{2,3,4\} \tag{42}$$

$$\mathcal{B}\backslash 1 = \{0,2,3,4\}. \tag{43}$$

Equations (40) and (41) reduce the computations of β_0^1 and β_1^1 to those of $F(\mathcal{A}\backslash 1,\mathcal{B})$ and $F(\mathcal{A}\backslash 1,\mathcal{B}\backslash 1)$.

Since, in this example, $\mathcal{A}\backslash 1 = \{2,3,4\}$ and $\mathcal{B}\backslash 1 = \{0,2,3,4\}$, the computation of $F(\mathcal{A}\backslash 1,\mathcal{B}\backslash 1)$ may be performed step by step using (33) as shown through the directed tree in Fig. 2. In Fig. 2, the recursive procedure for computing $F(\{2,3,4\}, \{0,2,3,4\})$ is described by a directed tree. In this tree, the P_j^ts are the weights associated with the edges. The plus sign, +, at a node represents the summation operation. The F function at a node labelled with a plus sign is obtained from the weighted summation of the F functions at the sons (defined to be nodes which have edges incident from the node under consideration) of the node, as seen from (33). Let a leaf denote a node which does not have sons. Each F at a leaf of the tree is directly computed using the initial condition given in (34). In the computation of $F(\{2,3,4\},\{0,2,3,4\})$, in this example, the nodes in the tree shown in Fig. 2 are visited via DFS. A node is not visited if the only branch incident on it has zero weight. As a result, the sons of this node are also not visited. In Fig. 2, the nodes having only incident dashed branches are not visited. In other words, the dashed branches in the tree are pruned because their weights are zero.

$F(\mathcal{A}\backslash 1,\mathcal{B})$ may be computed in a similar way as discussed above. Generally speaking, the computation in the recursive algorithm developed here is performed from top to bottom (or from left to right as shown in Fig. 2). In the average case scenario, a lot of branches in a directed tree have zero weight, such as the one shown in Fig. 2. The computational cost can be saved if a directed tree is visited from top to bottom since the nodes which have branches with zero weights incident onto them are not visited. In the worst case scenario, however, no branches have zero weight. The overall computation may be reduced if the nodes having the same F are merged so that they are visited only once. As shown in Fig. 2, almost every F appears twice at the leaves of the tree if the F functions at the nodes which have branches with zero weights incident on to them are computed. If the computation is performed from bottom to top, it is obvious that the computational cost can be reduced in the worst case scenario. However, in the average case scenario, the majority of the Fs at the leaves of a tree appears only once after the dashed branches are pruned, as shown in Fig. 2. Therefore, a lot of the computation might be wasted if the computation is performed from bottom to top since many nodes which are connected with dashed branches do not contribute to the F at the root of a tree. It is shown in the next section that the computational cost is lower if the computation is perform from top to bottom in the average case. The computation in the fast JPDA algorithm [12] may be interpreted as being performed from bottom to top if the procedure is organized into the same tree structure as discussed above. Detailed comparison between the recursive algorithm developed here and the fast JPDA algorithm will be given below.

V. COMPUTATIONAL COMPLEXITY ANALYSIS

The computational complexity measure here is defined in terms of the numbers of multiplications, $M(n,m)$, and additions, $A(n,m)$, used for calculating the β_j^ts in the JPDAF, where n is the number of targets and m is the number of measurements. Since the computational cost for normalizing β_j^t is negligible when compared with that of $F(\mathcal{A}\backslash t,\mathcal{B}\backslash j)$, multiplications and additions for normalizing β_j^t are not included in $M(n,m)$ and $A(n,m)$. In the worst case, the m measurements received fall inside the intersection of the validation gates of the n targets. Therefore, each measurement may be associated with either any one of the n targets or clutter. In the average case, however, it is assumed that there are on average two to three measurements which fall inside the validation gate of each target. In the following, the worst case analysis is given first. Then, the average case analysis is discussed.

A. Worst Case Analysis

In the worst case, no branch is pruned if the computation process is represented in a tree structure as shown in Fig. 2, without any dashed edges. As discussed in Section III, to compute all β_j^t,

only $F(\mathcal{A}\backslash t,\mathcal{B}\backslash j)$ for $t = 1,\ldots,n$, and $j = 1,\ldots,m$, and $F(\mathcal{A}\backslash 1,\mathcal{B})$ need to be computed. Therefore, $M(n,m)$ and $A(n,m)$ are, respectively, the number of multiplications and additions which suffice for computing nm β_j^ts ($j \neq 0$) and one β_0^t. In the recursive algorithm proposed in the previous section, the computation of the β_j^ts is based on (23) and (28). As shown there, the only operation required is one multiplication for computing each β_j^t after the corresponding F is computed. Therefore,

$$M(n,m) = nmM'(n-1,m) + M'(n-1,m+1) + nm + 1 \quad (44)$$

and

$$A(n,m) = nmA'(n-1,m) + A'(n-1,m+1) \quad (45)$$

where $M'(p,q)$ and $A'(p,q)$ denote, respectively, the numbers of multiplications and additions which suffice for computing the function $F(\mathcal{A},\mathcal{B})$ with $p = |\mathcal{A}|$ and $q = |\mathcal{B}|$.

In (33), there are q multiplications and $q-1$ additions used in the computation of $F(\mathcal{A},\mathcal{B})$ after $(q-1)$ $F(\mathcal{A}\backslash t,\mathcal{B}\backslash j)$s ($j \neq 0$) and one $F(\mathcal{A}\backslash t,\mathcal{B})$ are computed. Therefore,

$$M'(p,q) = (q-1)M'(p-1,q-1) + M'(p-1,q) + q \quad (46)$$

$$A'(p,q) = (q-1)A'(p-1,q-1) + A'(p-1,q) + q - 1 \quad (47)$$

with boundary conditions,

$$M'(1,q) = 0 \quad (48)$$

$$M'(p,2) = p-1, \quad \text{for} \quad p \geq 2 \quad (49)$$

$$M'(p,1) = p-1 \quad (50)$$

$$A'(1,q) = q-1 \quad (51)$$

$$A'(p,2) = p, \quad \text{for} \quad p \geq 2 \quad (52)$$

$$A'(p,1) = 0. \quad (53)$$

The boundary conditions for (46) and (47) are obtained from (34), (35), and (36).

In Table I, some sample values of $M(n,m)$ and $A(n,m)$ are provided. Comparing the listed values in Table I with the corresponding entries in Table II, it is not surprising to find that the recursive algorithm requires more multiplications and additions to compute the β_j^ts than the fast JPDA algorithm [12] in the worst case. This is because some Fs are computed more than once in the recursive algorithm as discussed in the previous section. However, the recursive algorithm is expected to perform much better than the fast JPDA algorithm in the average case. More discussion on the average case analysis will be given in the sequel.

TABLE I
Sample Values of $M(n,m)/A(n,m)$ in the Recursive Algorithm

n	m	$M(n,m)/A(n,m)$	n	m	$M(n,m)/A(n,m)$
3	6	134/582	5	10	43644/283900
4	8	2195/11952	6	12	1034045/7807860

TABLE II
Sample Values of $M(n,m)/A(n,m)$ in Fast JPDA Algorithm

n	m	$M(n,m)/A(n,m)$	n	m	$M(n,m)/A(n,m)$
3	6	150/396	5	10	10570/22100
4	8	1404/3232	6	12	70044/136584

TABLE III
Sample Values of $M(n,m)/A(n,m)$ in DFS Algorithm

n	m	$M(n,m)/A(n,m)$	n	m	$M(n,m)/A(n,m)$
3	6	318/787	5	10	96890/339041
4	8	5072/14849	6	12	2218992/9081085

In the worst case, the comparison between the computational complexity of the recursive algorithm and the DFS algorithm is also based on $M(n,m)$ and $A(n,m)$. In Table III, some sample values of $M(n,m)$ and $A(n,m)$ required for implementing the DFS algorithm are listed. Comparing Table III with Table I, it can be inferred that the $M(n,m)$s in the DFS algorithm are more than twice as large as those in the recursive algorithm and that the $A(n,m)$s in the DFS algorithm are always larger than those in the recursive algorithm. Considering the fact that the recursive algorithm is suitable for implementation in a multiprocessor system, the computational cost of the recursive algorithm could be much less than that of the DFS algorithm in the worst case.

The computational complexity given above is analyzed in terms of operational counts. It can also be given in "big O" notation. If the algorithm developed here is implemented in a system with a floating point unit, multiplication and addition operations would take about the same amount of time. Let $G(n,m) = M(n,m) + A(n,m)$. According to the definition given in [16, p. 2]AVA, the computational complexity of our algorithm in the worst case scenario is $O(G(n,m))$.

B. Average Case Analysis

In order to simplify the discussion, it is assumed that there are at most three measurements inside the validation gate of each target. This number was also selected in the examples given in [2, 8]. As a result, there are at most $3n$ β_j^ts ($j \neq 0$) and one β_0^t which need be computed. After the interference part of each β_j^t is obtained, $3n+1$ multiplications are sufficient for computing the β_j^ts as evident from (23) and (28).

TABLE IV
Upper Bounds of $M(n,m)$ and $A(n,m)$ in the Recursive Algorithm

n	m	$M(n,m)/A(n,m)$	n	m	$M(n,m)/A(n,m)$
3	-	50/150	5	-	1360/4080
4	-	273/819	6	-	6479/19437

Therefore, (44) and (45) lead to the inequalities,

$$M(n,m) \le 3nM'(n-1,m) + M'(n-1,m+1) + 3n + 1 \quad (54)$$

$$A(n,m) \le 3nA'(n-1,m) + A'(n-1,m+1). \quad (55)$$

In the average case, there are at most 4 multiplications and 3 additions required in the computation of $F(\mathcal{A},\mathcal{B})$ after 3 or less $F(\mathcal{A}\backslash t,\mathcal{B}\backslash j)$'s ($j \neq 0$) and one $F(\mathcal{A}\backslash t,\mathcal{B})$ are computed in (33). Again, from (46) and (47), we infer that

$$M'(p,q) \le 3M'(p-1,q-1) + M'(p-1,q) + 4 \quad (56)$$

$$A'(p,q) \le 3A'(p-1,q-1) + A'(p-1,q) + 3. \quad (57)$$

Since, in a cluttered environment, m is likely to be larger than n in the average case, therefore by using (48) and (51), the initial conditions for (56) and (57) may be simplified to

$$M'(1,q) = 0 \quad (58)$$

$$A'(1,q) \le 3. \quad (59)$$

With the above initial conditions, it is not difficult to show that both $M'(p,q)$ and $A'(p,q)$ are independent of q and that those are bounded from above by

$$M'(p,q) \le \tfrac{4}{3}(4^{p-1} - 1) \quad (60)$$

$$A'(p,q) \le 4^p - 1. \quad (61)$$

Substituting (60) into (54) and (61) into (55), the upper bounds for $M(n,m)$ and $A(n,m)$ in the average case can be obtained after routine manipulations.

$$M(n,m) \le \frac{3n+1}{3}(4^{n-1} - 1) \quad (62)$$

$$A(n,m) \le (3n+1)(4^{n-1} - 1). \quad (63)$$

Since the design of the fast JPDA algorithm [12] is based on the worst case scenario, there is no computational cost reduction in the average case. As a result, the recursive algorithm developed here is much more efficient than the fast JPDA algorithm, as shown in Table II and IV. If we consider the fact that the entries in Table IV are loose upper bounds, the actual computational cost of the recursive algorithm would be much less.

The upper bounds of $M(n,m)$ and $A(n,m)$ in the DFS algorithm, given in [13, 14], are reproduced below for ready reference

$$M(n,m) < 2*4^n - 2 - 3n - 3^n \quad (64)$$

$$A(n,m) < 4^n + 3n*4^{n-1}. \quad (65)$$

TABLE V
Upper Bounds of $M(n,m)/A(n,m)$ in DFS Algorithm

n	m	$M(n,m)/A(n,m)$	n	m	$M(n,m)/A(n,m)$
3	-	90/208	5	-	1788/4864
4	-	417/1024	6	-	7443/22528

Comparing the upper bounds of $M(n,m)$ in (64) and (62), it is apparent that the recursive algorithm requires less multiplications than the DFS algorithm for moderate value of n. A similar conclusion may be reached by comparing the upper bounds of $A(n,m)$ in (63) and (65). Furthermore, we infer that the recursive algorithm always requires less additions than the DFS algorithm. The computational complexity of the recursive algorithm and the DFS algorithm in the average case may be assessed by comparing sample values of the upper bounds of $M(n,m)$ and $A(n,m)$ given in Table IV for the recursive algorithm and in Table V for the DFS algorithm.

The computational complexity of our algorithm in the average case can also be given in terms of the "big O" notation. From (62) and (63), one can conclude that the computational complexity of our algorithm is $O(n4^n)$.

VI. CONCLUDING REMARKS

In this paper, a recursive algorithm, which is suitable for implementation in a multiprocessor system, is developed. In this algorithm, the computation of the *a posteriori* probabilities, β_j^ts, is not based on the generation of the data association hypotheses like in the DFS algorithm [13]. Each β_j^t in this algorithm is decomposed into two parts. The computation of one part is trivial and the recursive algorithm, developed here, is used to compute the other part (due to interference from other targets).

In the new algorithm, the computation of the interference part of β_j^t is implemented recursively in a top-to-bottom mode. In the worst case, this recursive algorithm requires more multiplications and additions than the fast JPDA algorithm [12]. However, in the average case, the recursive algorithm is expected to outperform the fast JPDA algorithm. In comparison with the DFS algorithm developed in [13, 14] the recursive algorithm requires less multiplications and additions than the DFS algorithm. The most important feature of the recursive algorithm is that it can be implemented on a multiprocessor system. Some suggestions for the implementation of the new algorithm on a multiprocessor system are given in Appendix A.

The task of computing $F(\mathcal{A}\backslash t, \mathcal{B}\backslash j)$ is related to that of evaluating the permanents, defined in [17], of an $m \times (n-1)$ matrix and its submatrices having P_l^τ for entries, where $\tau = 1,\ldots,n$, $\tau \neq t$, $l = 0,1,\ldots,m$, and $l \neq j$. Justifications are provided in Appendix B, where illustrative examples serve to verify the results obtained by exploiting this relationship.

APPENDIX A: SUGGESTIONS FOR IMPLEMENTATION

In order to explore the inherited parallelism in our efficient algorithm, the computation of the *a posteriori* probabilities $\beta_j^t(\mathcal{A},\mathcal{B})$ in our algorithm may be summarized as below.

1) For $j = 1,2,\ldots,m$, $t = 1,2,\ldots,n$, and $P_j^t \neq 0$,

$$\beta_j^t(\mathcal{A},\mathcal{B}) = P_j^t F(\mathcal{A}\backslash t, \mathcal{B}\backslash j) \tag{66}$$

and for $t = 1,2,\ldots,n$,

$$\beta_0^t = P_0 F(\mathcal{A}\backslash t, \mathcal{B}). \tag{67}$$

2) For any $\mathcal{A}$ and $\mathcal{B}$,

$$F(\mathcal{A},\mathcal{B}) = P_0 F(\mathcal{A}\backslash t, \mathcal{B}) + \sum_{j \in \mathcal{B}\backslash 0} P_j^t F(\mathcal{A}\backslash t, \mathcal{B}\backslash j). \tag{68}$$

3) The constant

$$c = \sum_{j=0}^{m} \beta_j^t(\mathcal{A},\mathcal{B}) \tag{69}$$

may be computed for any value of t as done in (29) for the $t = 1$ case.

4) For $j = 0,1,\ldots,m$, $t = 1,2,\ldots,n$, and $\beta_j^t \neq 0$,

$$\beta_j^t \leftarrow \frac{1}{c}\beta_j^t. \tag{70}$$

To implement our efficient algorithm, memory space has to be allocated for β_j^t, P_j^t, $\mathcal{A}$, and $\mathcal{B}$. Let BETA$(0:m, 1:n)$ and $P(0:m, 1:n)$ be two 2-D arrays each of size $(m+1) \times n$ for storing β_j^t and P_j^t. Since $\mathcal{A}$ and $\mathcal{B}$ are ordered sets of integers, some special data structures may be required to stored these two sets. As shown in Fig. 2, two types of operations are required on $\mathcal{A}$ and $\mathcal{B}$ to compute of $F(\mathcal{A},\mathcal{B})$. One operator deletes an element from either $\mathcal{A}$ or $\mathcal{B}$ when the computation advances from a node at the higher level of the tree to a node at the lower level. The other operator inserts an element in either $\mathcal{A}$ or $\mathcal{B}$ when the computation backtracks from a node at the lower level of the tree to a node at the higher level. Thus each operation updates either $\mathcal{A}$ or $\mathcal{B}$. In order to implement the two operations efficiently, data structures can be found in [16] for storing $\mathcal{A}$ and $\mathcal{B}$. Let the two functions, Delete$(\mathcal{X}, j)$ and Insert$(\mathcal{X}, j)$, denote, respectively, the operations of deletion of element j from set $\mathcal{X}$ and insertion of element j into set $\mathcal{X}$ when $j \neq 0$. In the case when $j = 0$, Delete and Insert do nothing. To compute $F(\mathcal{A},\mathcal{B})$, a function can be defined using (A.3). For the sake of simplicity, let $F(\mathcal{A},\mathcal{B})$ be that function which returns the value on the right-hand side of (68).

With the notations introduced above, the computation of *a posteriori* probabilities $\beta_j^t(\mathcal{A},\mathcal{B})$ in our efficient algorithm for implementation on a multiprocessor system is summarized below.

1) For $j = 0,1,2,\ldots,m$, $t = 1,2,\ldots,n$, and $P(j,t) \neq 0$,

$$\text{BETA}(j,t) = P(j,t)F(\text{Delete}(\mathcal{A},t), \text{Delete}(\mathcal{B},j)).$$

2) The normalization constant for any target t is

$$c = \sum_{j=0}^{m} \text{BETA}(j,t).$$

3) For $j = 0,1,2,\ldots,m$, $t = 1,2,\ldots,n$, and BETA$(j,t) \neq 0$,

$$\text{BETA}(j,t) \leftarrow \frac{1}{c}\text{BETA}(j,t).$$

Each BETA(j,t) in steps 1 and 3 above can be computed in parallel in a multiprocessor system such as CM-200 from Connection Machine Inc.. A programmer may view the CM-200 as a set of *virtual processors*, one for each data element. The corresponding code and data are passed to each processor. On a CM-200, steps 1 and 3 may be implemented using a conditional FORALL statement in CM FORTRAN [18]. The effect of a conditional FORALL statement is similar to the embedding of an IF statement in a DO loop using FORTRAN 77. With a conditional FORALL statement, all data elements are operated on simultaneously. In general, our efficient algorithm is ready to be implemented in any parallel computer with single-instruction and multiple-data (SIMD) or multiple-instruction and multiple-data (MIMD) architecture.

APPENDIX B: RELATIONSHIP BETWEEN $F(\mathcal{A},\mathcal{B})$ AND PERMANENT

The definition for a permanent is given below [17].

DEFINITION Let $A = (a_j^t)$ be an $m \times n$ matrix over any commutative ring, $m \leq n$. The *permanent* of A, written Per(A) is defined by

$$\text{Per}(A) = \sum_{\sigma} a_1^{\sigma(1)} a_2^{\sigma(2)} \cdots a_m^{\sigma(m)} \tag{71}$$

where the summation extends over each of the $n!/(n-m)!$ one-to-one maps, denoted by σ, of the set $\{1,2,\ldots,m\}$ to the set $\{1,2,\ldots,n\}$.

In our case, let $\mathbf{P} = (P_j^t)$ denote the matrix having $(m+1)$ rows and n columns. $\mathcal{A}$ and $\mathcal{B}$, defined before, contain, respectively, the column and row indices of P, i.e., $\mathcal{A} = \{1,2,\ldots,n\}$ and $\mathcal{B} = \{0,1,\ldots,m\}$. For notational convenience, in later usage, we denote the permanent of P, Per(P), by $\mathrm{Per}(\mathcal{A},\mathcal{B})$. This causes no confusion or ambiguity because the 2-tuple (j,t) formed from an element, t, of $\mathcal{A}$ and an element, j, of $\mathcal{B}$ uniquely defines P_j^t. Therefore, applying the definition for permanent to a matrix or its transpose, we get

$$\mathrm{Per}(\mathcal{A},\mathcal{B}) = \sum_{\sigma} P_{j_1}^{\sigma(j_1)} P_{j_2}^{\sigma(j_2)} \cdots P_{j_{|\mathcal{B}|}}^{\sigma(j_{|\mathcal{B}|})}, \quad |\mathcal{B}| \le |\mathcal{A}| \tag{72}$$

$$\mathrm{Per}(\mathcal{A},\mathcal{B}) = \sum_{\sigma} P_{\sigma(t_1)}^{t_1} P_{\sigma(t_2)}^{t_2} \cdots P_{\sigma(t_{|\mathcal{A}|})}^{t_{|\mathcal{A}|}}, \quad |\mathcal{B}| > |\mathcal{A}|. \tag{73}$$

It is not difficult to show that $F(\mathcal{A},\mathcal{B})$ in (33) is expandable as

$$F(\mathcal{A},\mathcal{B}) = (P_0)^{|\mathcal{A}|} + (P_0)^{|\mathcal{A}|-1}\left(\sum_{t\in\mathcal{A}} \mathrm{Per}(\{t\},\mathcal{B}\backslash 0)\right) + (P_0)^{|\mathcal{A}|-2}\left(\sum_{t_1,t_2\in\mathcal{A}} \mathrm{Per}(\{t_1,t_2\},\mathcal{B}\backslash 0)\right) + \cdots + (P_0)^{|\mathcal{A}|-|\mathcal{B}\backslash 0|}\left(\sum_{t_1,\ldots,t_{|\mathcal{B}\backslash 0|}\in\mathcal{A}} \mathrm{Per}(\{t_1,\ldots,t_{|\mathcal{B}\backslash 0|}\},\mathcal{B}\backslash 0)\right), \quad |\mathcal{B}\backslash 0| \le |\mathcal{A}|, \tag{74}$$

$$F(\mathcal{A},\mathcal{B}) = (P_0)^{|\mathcal{A}|} + (P_0)^{|\mathcal{A}|-1}\left(\sum_{t\in\mathcal{A}} \mathrm{Per}(\{t\},\mathcal{B}\backslash 0)\right) + (P_0)^{|\mathcal{A}|-2}\left(\sum_{t_1,t_2\in\mathcal{A}} \mathrm{Per}(\{t_1,t_2\},\mathcal{B}\backslash 0)\right) + \cdots + \mathrm{Per}(\mathcal{A},\mathcal{B}\backslash 0) \quad |\mathcal{B}\backslash 0| > |\mathcal{A}|. \tag{75}$$

The validities of (74) and (75) are illustrated by the consideration of some special cases below.

EXAMPLE 1 Let $\mathcal{A} = \{t\}$, so that $|\mathcal{A}| = 1$. In this case, the use of (74) or (75), as the case demands, yields

$$F(\mathcal{A},\mathcal{B}) = P_0 + \mathrm{Per}(\mathcal{A},\mathcal{B}\backslash 0) = P_0 + \sum_{j\in\mathcal{B}\backslash 0} P_j^t = \sum_{j\in\mathcal{B}} P_j^t. \tag{76}$$

The above equation agrees with (34).

EXAMPLE 2 Let $\mathcal{B} = \{0\}$, so that $|\mathcal{B}| = 1$ and $|\mathcal{B}\backslash 0| = 0$. In this case, $|\mathcal{A}| \ge 1$ and (74) applies, yielding

$$F(\mathcal{A},\mathcal{B}) = (P_0)^{|\mathcal{A}|}. \tag{77}$$

The above equation agrees with (35).

EXAMPLE 3 Let $\mathcal{B} = \{0,j\}$, where $j \ne 0$, and $|\mathcal{A}| \ge 2$, so that $|\mathcal{B}| = 2$ and $|\mathcal{A}| > |\mathcal{B}\backslash 0|$. Then (74) applies in this case, to yield equation (36).

EXAMPLE 4 Let $\mathcal{A} = \{1,3,4\}$ and $\mathcal{B} = \{0,2,4,5,7\}$. The matrix (P_j^t), where the indices t and j are drawn from the sets $\mathcal{A}$ and $\mathcal{B}$, is (note that $P_0^t \triangleq P_0$)

$$\begin{pmatrix} P_0 & P_0 & P_0 \\ P_2^1 & P_2^3 & P_2^4 \\ P_4^1 & P_4^3 & P_4^4 \\ P_5^1 & P_5^3 & P_5^4 \\ P_7^1 & P_7^3 & P_7^4 \end{pmatrix}.$$

In this case, $|\mathcal{A}| < |\mathcal{B}\backslash 0|$. Then (75) applies, yielding

$$F(\mathcal{A},\mathcal{B}) = (P_0)^3 + (P_0)^2 \sum_{t\in\mathcal{A}} \sum_{j\in\mathcal{B}\backslash 0} P_j^t + P_0 \sum_{\substack{t_1,t_2\in\mathcal{A} \\ t_1\ne t_2}} \sum_{\substack{j_1,j_2\in\mathcal{B}\backslash 0 \\ j_1\ne j_2}} P_{j_1}^{t_1} P_{j_2}^{t_2} + \sum_{\substack{j_1,j_2,j_3\in\mathcal{B}\backslash 0 \\ j_1\ne j_2\ne j_3}} P_{j_1}^1 P_{j_2}^3 P_{j_3}^4. \tag{78}$$

$F(\mathcal{A},\mathcal{B})$ can also be calculated as below, using (33)

$$F(\mathcal{A},\mathcal{B}) = (P_0)^2 \sum_{j\in\mathcal{B}} P_j^4 + P_0 \sum_{j_1\in\mathcal{B}\backslash 0} P_{j_1}^3 \sum_{j_2\in\mathcal{B}\backslash j_1} P_{j_2}^4 + P_0 \sum_{j_1\in\mathcal{B}\backslash 0} P_{j_1}^1 \sum_{j_2\in\mathcal{B}\backslash j_1} P_{j_2}^4 + \sum_{j_1\in\mathcal{B}\backslash 0} P_{j_1}^1 \sum_{j_2\in\mathcal{B}\backslash\{0,j_1\}} P_{j_2}^3 \sum_{j_3\in\mathcal{B}\backslash\{j_1,j_2\}} P_{j_3}^4. \tag{79}$$

It is not difficult to show that (78) and (79) are equivalent.

ACKNOWLEDGMENT

The authors thank Jongtae Chun for his careful reading of the manuscript.

REFERENCES

[1] Bar-Shalom, Y., and Fortmann, T. E. (1988) *Tracking and Data Association.* Orlando, FL: Academic Press, 1988.

[2] Fortmann, T. E., Bar-Shalom, Y., and Scheffe, M. (1983) Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, **OE-8** (July 1983), 173–184.

[3] Reid, D. B. (1979) An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, **AC-24** (Dec. 1979), 843–854.

[4] Kurien, T., and Washburn, R. B., Jr. (1985)
Multiobject tracking using passive sensors.
In *Proceedings of the American Controls Conference*, Boston, MA, June 1985, 1032–1038.

[5] Kurien, T., and Liggins, M. E., II (1988)
Report-to-track assignment in multisensor multitarget tracking.
In *Proceedings of the 27th Conference on Decision and Control*, Austin, TX, Dec. 1988, 2484–2488.

[6] Bar-Shalom, Y. (1990)
Multitarget-Multisensors Tracking: Advanced Applications.
Norwood, MA: Artech House, 1990.

[7] Fitzgerald, R. J. (1986)
Development of practical PDA logic for multitarget tracking by microprocessor.
In *Proceedings of the American Controls Conference*, Seattle, WA, June 1986, 889–898.

[8] Sengupta, D., and Iltis, R. A. (1989)
Neural solution to the multiple target tracking data association problem.
IEEE Transactions on Aerospace and Electronic Systems, **25** (Jan. 1989), 96–108.

[9] Sengupta, D., and Iltis, R. A. (1990)
Multiple maneuvering target tracking using neural networks.
Technical Report 90-32, Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, Dec. 1990.

[10] Zhou, B., and Bose, N. K. (1993)
A comprehensive analysis of "neural solution to the multitarget tracking data association problem".
IEEE Transactions on Aerospace and Electronic Systems, **29** (Jan. 1993), 260–263.

[11] Nagarajan, V., Chidambara, M. R., and Sharma, R. N. (1987)
Combinatorial problems in multitarget tracking—A comprehensive solution.
IEE Proceedings, Pt. F, *Communications, Radar and Signal Processing*, **134** (Feb. 1987), 113–118.

[12] Fisher, J. L., and Casasent, D. P. (1989)
Fast JPDA multitarget tracking algorithm.
Applied Optics, **28** (Jan. 1989), 371–376.

[13] Zhou, B., and Bose, N. K. (1993)
Multitarget tracking in clutter: Fast algorithms for data association.
IEEE Transactions on Aerospace and Electronic Systems, **29** (Apr. 1993), 352–363.

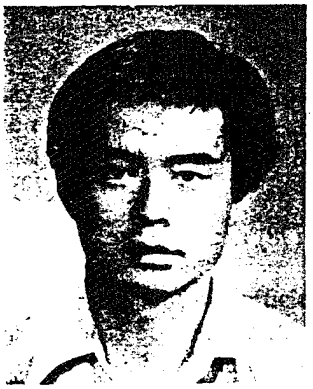
[14] Zhou, B. (1992)
Multitarget tracking in clutter: Algorithms for data association and state estimation.
Ph.D. dissertation, Pennsylvania State University, Department of Electrical and Computer Engineering, University Park, PA 16802, May 1992.

[15] Bar-Shalom, Y., and Tse, E. (1975)
Tracking in a cluttered environment with probabilistic data association.
Automatica, **11** (Sept. 1975), 451–460.

[16] Aho, A. V., Hopcroft, J. E., and Ullman, J. D. (1974)
The Design and Analysis of Computer Algorithms.
Reading, MA: Addison-Wesley, 1974.

[17] Minc, H. (1978)
Permanents.
Reading, MA: Addison-Wesley, 1978.

[18] Thinking Machines Corp. (1991)
Getting Started in CM Fortran.
Cambridge, MA: Thinking Machines Corp., Nov. 1991.



Bin Zhou (S'86—M'93) received the B.S. degree in electrical engineering from Fudan University, Shanghai, China, in 1982, the M.S. degree in computer engineering from Shanghai Institute of Technical Physics, China, in 1985, and the Ph.D. degree in electrical engineering from Pennsylvania State University, University Park, in 1992.

Dr. Zhou was a postdoctoral scholar at the Center for Spatial and Temporal Signal Processing and the Center for Multivariate Analysis, Pennsylvania State University. He is now with the Johnson School of Management at Cornell University. His research interests are in the areas of multitarget tracking, neural networks, and signal processing.

N. K. Bose (S'66—M'67—SM'74—F'81) received his B.Tech (with honors), M.S., and Ph.D. degrees, all in electrical engineering, at Indian Institute of Technology, Kharagpur, India, Cornell University, Ithaca, NY, and Syracuse University, Syracuse, NY, respectively.

He is currently the HRB-Systems Professor and Director of the Spatial and Temporal Signal Processing Center at Pennsylvania State University, University Park, PA.

Dr. Bose is the author or editor of several books and numerous papers. He is currently Editor-in-Chief of *Multidimensional Systems and Signal Processing*.

Pergamon

0005-1098(93)E0042-3

Automatica, Vol. 30, No. 9, pp. 1469–1472, 1994

0005-1098/94 $7.00 + 0.00

Brief Paper

A Unified Approach to Data Association in Multitarget Tracking*†

B. ZHOU‡ and N. K. BOSE‡

Key Words—Target tracking; tracking systems.

Abstract—A unified framework is proposed to provide a comprehensive understanding of the problem of data association in multitarget tracking. Under this framework, a systematic scheme is developed for generating the data association hypotheses in the target-oriented, measurement-oriented, and track-oriented approaches. Since there are many data association hypotheses with identical likelihoods in the measurement-oriented and track-oriented approaches, two specialized algorithms are developed to efficiently generate the data association hypotheses in the two approaches by adapting the depth-first search (DFS) algorithm developed earlier in the target-oriented case.

1. *Introduction*

RECENTLY, considerable attention has been paid to the problem of data association in multitarget tracking. In a cluttered environment, the detected signals comprising the measurements may not all originate from the targets of interest. Some of those could be either from clutter or false alarm. Each detected signal may be used as a measurement to update the state of a target. Therefore, more than one measurement may be available for the purpose of target state updating. The ambiguity occurring in associating previously known targets with new measurements is referred to as the problem of data association. To solve this problem, three approaches have been reported in the literature. Those are the *target-oriented, measurement-oriented,* and *track-oriented* approaches. In the target-oriented approach, each measurement is assumed to have originated from either a known target or clutter (Bar-Shalom and Fortmann, 1988; Bar-Shalom and Tse, 1975; Fortmann *et al.* 1983). In the measurement-oriented approach, each measurement is hypothesized to have originated from either a known target, a new target, or clutter (Reid, 1979). In the track-oriented approach, however, each track is hypothesized to be either undetected, terminated, associated with a measurement, or linked to the start of a maneuver (Kurien and Washburn, 1985; Kurien and Liggins, 1988; Bar-Shalom, 1990). In multitarget tracking, Fortmann *et al.* (1983) proposed the Joint Probabilistic Data Association Filter (JPDAF) as an implementation of the target-oriented approach. Since the JPDAF, which may be viewed to be synonymous with the target-oriented approach, lacks track initiation capability and also suffers from the track biases and coalescence problem (Bar-Shalom, 1990; Fitzgerald, 1986), further study of the measurement-oriented and the track-oriented approaches becomes necessary to address the computation intensive algorithms associated with these approaches in comparison to the JPDAF. Let n and m denote, respectively, the number of targets and the number of measurements. In the worst case, each of the n targets could be associated with any one of the m measurements. It has been shown that the computational cost for associating the n targets with the m measurements increases exponentially as a function of the parameters n and m (Zhou and Bose, 1993). However, if the data association problem involving n targets and m measurements can be decomposed into smaller subproblems, then the overall computational cost may, approximately, become polynomially dependent on n and m. The resulting small groups of targets and measurements are referred to as clusters in the multitarget tracking terminology. Two targets are in the same cluster if and only if at least one measurement falls inside the intersection of the validation gates of the two targets. A validation gate of a target is a region of certain size that surrounds the predicted position of the target. In this paper, the data association problem is with reference to measurements and targets belonging to the same cluster.

After targets are properly clustered, Zhou and Bose (1993) proposed a depth-first search (DFS) algorithm to efficiently generate the data association hypotheses in the JPDAF. They showed that a significant reduction in the computational cost of implementation of the JPDAF can be realized by the sharing of common factors occurring in the calculation of the conditional probabilities of the data association hypotheses. In this paper, the problems of generation of data association hypotheses in the measurement-oriented and track-oriented approaches are reformulated to permit development of a mathematical model, similar to the one developed by Zhou and Bose (1993) for the data association problems in the target-oriented case. A unified framework is provided for adapting the DFS algorithm to efficiently generate the data association hypotheses in the measurement-oriented and track-oriented approaches, where the total numbers of data association hypotheses are expected to increase drastically over the target-oriented approach. However, reduction in the overall computational cost may be feasible from observations on the conditional likelihood of a data association hypothesis. In the target-oriented approach, the conditional likelihood of each data association hypothesis is unique. When targets are grouped into clusters, this uniqueness property does not hold for the measurement-oriented and track-oriented approaches. Our goal here is to develop two specialized DFS algorithms which can efficiently identify the data association hypotheses with identical conditional likelihood in the measurement-oriented and track-oriented approaches.

2. *Problem formulation*

Suppose that there are two targets moving towards each other. In one radar scan, three measurements are received, as shown in Fig. 1. In the target-oriented approach, a measurement may originate from either a known target or clutter. Zhou and Bose (1993) described this situation using three finite sets of ordered integers, Z_j $(j = 1, 2, 3)$. The three

* Received 19 March 1993; revised 21 July 1993; received in final form 10 October 1993. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Editor A. P. Sage. Corresponding author Professor N. K. Bose. Fax 814 865 7065.

† This research was supported by SDIO/IST under ONR contract N00014-92-J-1755.

‡ The Spatial and Temporal Signal Processing Center, Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802, U.S.A.

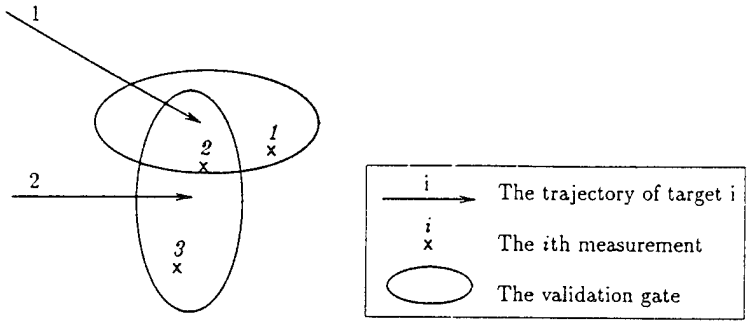


FIG. 1. A typical layout for the validation gates in the case of 2 targets and 3 measurements.

sets are

$$Z_1 = \{0, 1\}, \quad Z_2 = \{0, 1, 2\}, \quad Z_3 = \{0, 2\}.$$

$Z_1 = \{0, 1\}$ implies that measurement 1 could originate from either clutter or target 1. Similar explanations may be given to Z_2 and Z_3. In the measurement-oriented approach, however, any one of the three measurements could have originated from a new target. The three-finite sets of ordered integers may be augmented in the following manner to incorporate the possibility that a measurement may originate from a new target:

$$Z_1 = \{0, 1, 3\}, \quad Z_2 = \{0, 1, 2, 4\}, \quad Z_3 = \{0, 2, 5\}.$$

Note that $Z_1 = \{0, 1, 3\}$ implies that measurement 1 could originate either from clutter, or old target 1, or a new target 3. Similar comments apply to Z_2 and Z_3. In addition to the possibility that a measurement could originate from a new target, an established track could either be terminated or linked to the start of a maneuver in the track-oriented approach. For the sake of simplicity, suppose that Singer's dynamic model (Singer, 1970) is used. Since the maneuverability of a target is inherited in Singer's dynamic model, it is not necessary to consider it in data association. On the other hand, a new integer set can be created to allow for the possibility that a track may be terminated at any moment. For the example shown in Fig. 1, four sets of finite ordered integers may be used:

$$Z_0 = \{1, 2\}, \quad Z_1 = \{0, 1, 3\}, \quad Z_2 = \{0, 1, 2, 4\}, \quad Z_3 = \{0, 2, 5\},$$

where $Z_0 = \{1, 2\}$ implies that the tracks of targets 1 and 2 may be terminated. Note that, as in the measurement-oriented approach, $Z_2 = \{0, 1, 2, 4\}$ implies that measurement 2 could originate either from clutter, old target 1 or 2, or a new target 4.

By comparing the data association problems in the measurement-oriented and track-oriented approaches with their counterpart in the JPDAF, the DFS algorithm developed by Zhou and Bose (1993) may be easily extended to generate the data association hypotheses in the measurement and track-oriented approaches. Since a measurement or a track may be associated with more choices in these two approaches in contrast to the target-oriented case, the total number of data association hypotheses is expected to increase. Therefore, it is important to reduce the computational cost in such situations.

In the target-oriented approach, the conditional likelihood of each data association hypothesis is unique. This uniqueness property does not hold, in general, in the measurement-oriented and track-oriented approaches. Consider, again, the example described in Fig. 1. Suppose that measurements 1, 2 and 3 originate, respectively, from target 1, clutter, and a new target. Then, the conditional likelihood of this data association hypothesis is (Reid, 1979):

$$P = \frac{1}{c} P_{11} \lambda_C \lambda_{NT}, \tag{1}$$

where c is a constant, P_{11} is the likelihood that measurement 1 originates from target 1, λ_C is the density of clutter, and λ_{NT} is the density of previously unknown targets. Alternatively, measurement 2 could originate from a new target and measurement 3 could originate from clutter. It can be verified that the conditional likelihood of this data association hypothesis equals the one given in (1). Therefore, there are at least two data association hypotheses with equal conditional likelihood in this example. In general, λ_C and λ_{NT} may not be constant in a surveillance region. For example, clutter distribution may not be uniform and the measurements which lie in the center of the surveillance region (or field-of-view) are less likely to come from new targets than those which lie near the boundary of the surveillance region. However, measurements and targets are normally grouped into clusters in multitarget tracking. Within each cluster, λ_C and λ_{NT} can be treated as constants. When the number of targets and the number of measurements in a cluster are large, duplications in conditional likelihood can occur much more often. The same observation can be made in the track-oriented approach. Therefore, specialized DFS algorithms which can effectively identify such duplications in the measurement-oriented and track-oriented approaches need to be developed to reduce computational cost significantly.

3. *Development of an unified framework*

In the target-oriented approach, the received measurements are divided into two groups. The measurements in the first group are assumed to be associated with some known targets and the measurements in the second group are assumed to be from clutter. The data association between the known targets and the measurements in the first group is accomplished by the DFS algorithm (Zhou and Bose, 1993). The association between clutter and the measurements in the second group is trivial.

In the measurement-oriented approach, a measurement could be associated with either a known target or clutter or a new target. To accommodate these three possibilities, the received measurements may be divided into three groups. As pointed out in the previous section, some data association hypotheses in the measurement-oriented approach may have identical conditional likelihood. If those hypotheses can be identified, the corresponding likelihood can then be computed only once. Suppose that there are m measurements. In a particular data association hypothesis ε, N_T measurements are associated with previously known targets, N_C measurements are associated with clutter, and N_{NT} measurements are associated with new targets. Therefore, $m = N_T + N_C + N_{NT}$. It can be shown that a factor of the conditional likelihood of ε which corresponds to the association between N_C measurements and clutter is proportional to $\lambda_C^{N_C}$. Similarly, a factor of the conditional likelihood of ε which corresponds to the association between N_{NT} measurements and new targets is proportional to $\lambda_{NT}^{N_{NT}}$. When a new data association hypothesis is generated by merely shuffling the measurements associated with clutter in ε with those associated with new targets, the conditional likelihood of this new hypothesis will be the same as that of ε as long as N_C and N_{NT} are unchanged. This observation

suggests that the received measurements should be partitioned in a certain way for the purpose of identifying data association hypotheses with identical conditional likelihood.

The difference between the track-oriented and measurement-oriented approaches is in the manner of handling of the previously known targets which are not associated with any measurements in a data association hypothesis. In the measurement-oriented approach, such targets are simply classified as undetected. However, in the track-oriented approach, those targets could be either undetected or their trajectories could have been terminated. In other words, the previously known targets are divided into three distinct groups in the track-oriented approach. The partitioning of the previously known targets can be done in the same way as the partitioning of the measurements in the measurement-oriented approach.

In conclusion, the generation of the data association hypotheses in all three approaches may be described as below.

1. Select the measurements and the previously known targets which are to be associated with each other. If the target-oriented approach is used, then a data association hypothesis is generated and this process is repeated until all hypotheses are generated. Otherwise, go to step 2.
2. Partition the remaining measurements into two groups. The measurements in one group are hypothesized to be associated with clutter. The measurements in the other one are then to be associated with new targets. If the measurement-oriented approach is used, then go back to step 1 and this process is repeated until all data association hypotheses are generated. Otherwise, go to step 3.
3. Partition the remaining known targets into two groups. The targets in one group are hypothesized to be undetected. The targets in the other group have their trajectories terminated. Repeat steps 1–3 until all data association hypotheses are generated.

The above description provides a unified framework for the generation of the data association hypotheses in all three approaches. It can be claimed that the measurement-oriented approach is a special case of the track-oriented approach when no target has its trajectory terminated and the target-oriented approach is a special case of the measurement-oriented approach when there are no new targets. Under this unified framework, the association between the measurements and the previously known targets can be established by adapting the DFS algorithm originally developed for the target-oriented case (Zhou and Bose, 1993).

4. *Two DFS algorithms*

The data association problem in the measurement-oriented approach, as discussed in the last section, can be solved in two stages. First, the association between the measurements and the previously known targets can be established using the DFS algorithm (Zhou and Bose, 1993). Then, the remaining measurements are divided into two groups. The measurements in one group are assumed to be associated with clutter, and those in the other group are assumed to originate from new targets. Nijenhuis and Wilf (1975) presented a partitioning algorithm which can efficiently draw different samples of size k out of N objects. After N_T measurements are associated with N_T targets, N_C measurements may be selected from the $(m - N_T)$ remaining measurements for association with clutter. Subsequently, $N_{NT} = (m - N_T - N_C)$ measurements are assumed to be from new targets. By cascading the DFS algorithms and the partitioning algorithm, a new DFS algorithm may be developed to generate all data association hypotheses in the measurement-oriented approach. After each association between N_T measurements and N_T targets is established, for a given N_C $(= 0, \ldots, m - N_T)$, the likelihood $\lambda_{NT}^{m-N_T-N_C}\lambda_C^{N_C}P$ is computed only once. However, for a given N_C, $(m - N_T)!/(N_C!\,(m - N_T - N_C)!)$ different partitions can be generated, where $K!$ denotes the factorial of K. This means that the $(m - N_T)!/(N_C!(m - N_T - N_C)!)$ data association hypotheses with identical likelihood are effectively grouped together in this new DFS algorithm.

In the track-oriented approach, both the remaining measurements and the remaining targets are divided into two groups after N_T measurements are associated with N_T targets. Suppose that N_D targets have their tracks terminated (or discontinued) and N_{ND} targets are not detected. If there are n previously known targets, then $n = N_T + N_D + N_{ND}$. Similar to the measurement-oriented approach, a new DFS algorithm may be developed by cascading the DFS algorithm in the target-oriented case (Zhou and Bose, 1993) and two partitioning algorithms (Nijenhuis and Wilf, 1975). In the track-oriented approach, there could be

$$\frac{(m - N_T)!\,(n - N_T)!}{N_C!\,(m - N_T - N_C)!\,N_D!\,(m - N_T - N_D)!}$$

data approach hypotheses with identical likelihood.

As shown above, a lot of computation may be saved by using the two new DFS algorithms to generate data association hypotheses in the measurement-oriented and track-oriented approaches. The computational complexity of the two algorithms may be analyzed as was done for the DFS algorithm in the target-oriented case (Zhou and Bose, 1993).

5. *Conclusion and remarks*

The unified framework developed in this paper provides a systematic scheme for the generation of data association hypotheses in the target-oriented, measurement-oriented and track-oriented approaches. Furthermore, the framework makes the development of efficient algorithms possible for the measurement-oriented and track-oriented approaches by adapting the DFS algorithm developed in the target-oriented scenario. The success of the algorithms, in the measurement-oriented and track-oriented approaches, is derived from exploitation of the fact that, after the targets and measurements are grouped into clusters, many data association hypotheses share an identical conditional likelihood. The computational cost can be saved significantly by computing the identical likelihood only once in each such situation.

The ambiguity of data association in the measurement-oriented and track-oriented approaches may be further resolved using the data in the N subsequent scans. Singer *et al.* (1974) showed that near-optimal performance was achieved with $N = 1$ for the single target case. In the case of multitarget tracking, the data from two successive scans $(N = 2)$ were used in data association (Bar-Shalom, 1990). The process of using data from more than one scan in data association is called multiscan correlation. The application of the algorithms developed in this paper to multiscan correlation is straightforward. To generate data association hypotheses in multiscan correlation in the measurement-oriented and track-oriented approaches, the algorithms proposed in Section 4 may be applied to generate data association hypotheses in each scan.

Acknowledgements—The authors wish to acknowledge the constructive comments of the reviewers. This research was supported by SDIO/IST under ONR contract N00014-92-J-1755.

References

Bar-Shalom, Y. (1990). *Multitarget-multisensors Tracking: Advanced Applications*. Artech House Inc., Norwood, MA.

Bar-Shalom, Y. and T. E. Fortmann (1988). *Tracking and Data Association*. Academic Press, Orlando, FL.

Bar-Shalom, Y. and E. Tse (1975). Tracking in a cluttered environment with probabilistic data association. *Automatica*, **11**, 451–460.

Fitzgerald, R. J. (1986). Development of practical PDA logic for multitarget tracking by microprocessor. In *Proc. American Controls Conference*, Seattle, WA, pp. 889–898.

Fortmann, T. E., Y. Bar-Shalom and M. Schaffe (1983). Sonar tracking of multiple targets using joint probabilistic data association. *IEEE J. Oceanic Engng*, **OE-8**, 173–184.

Kurien, T. and M. E. Liggins II (1988). Report-to-track assignment in multisensor multitarget tracking. In *Proc. 27th Conf. on Decision and Control*, Austin, TX, pp. 2484–2488.

Kurien, T. and R. B. Washburn Jr (1985). Multiobject tracking using passive sensors In *Proc. American Controls Conference*, Boston, MA, pp. 1032–1038.

Nijenhuis, A. and H. S. Wolf (1975). *Combinatorial Algorithms*. Academic Press, Orlando, FL.

Reid, D. B. (1979). An algorithm for tracking multiple targets. *IEEE Trans. Automatic Control*, **AC-24**, 843–854.

Singer, R. A. (1970). Estimating optimal tracking filter performance for manned maneuvering targets. *IEEE Trans. Aerospace Electronic Systems*, **AES-6**, 473–483.

Singer, R. A., R. G. Sea and K. B. Housewright (1974). Derivation and evaluation of improved tracking filters for use in dense multitarget environments. *IEEE Trans. Information Theory*, **IT-20**, 423–432.

Zhou, B. and N. K. Bose (1993). Multitarget tracking in clutter: fast algorithms for data association. *IEEE Trans. Aerospace Electronic Systems*, **AES-29**, 352–363.

DIGITAL SIGNAL PROCESSING 4, 40–48 (1994)

A Novel Subspace-Based Approach to Parameter Estimation*

Jongtae Chun and N. K. Bose

The Spatial and Temporal Signal Processing Center, Electrical and Computer Engineering Department, The Pennsylvania State University, 121 Electrical Engineering East, University Park, PA 16802

Chun, J., and Bose, N. K., A Novel Subspace-Based Approach to Parameter Estimation, *Digital Signal Processing* 4 (1994), 40–48.

The known eigendecomposition-based subspace techniques for parameter estimation of multiple signals in additive white noise or the directions of arrivals of wavefronts at an array of sensors use either the information generated from the noise subspace or the signal subspace. Here, we directly use information from both the noise and signal subspaces to estimate the parameters. Comparisons with other subspace techniques are made and an illustrative example is provided.

1. INTRODUCTION

Since 1979, several subspace methods for parameter estimation have been introduced. MUSIC, ESPRIT, and their variants such as TLS-ESPRIT and GEESE, which can be applied to estimate the directions of arrivals of impinging wavefronts by operating on the signals collected at an array of sensors, are described in [1]. More recently, the weighted subspace method introduced in [2] has been shown to be asymptotically efficient on the estimation error variance, as the stochastic maximum likelihood technique is known to be, under certain regularity conditions involving the assumption of Gaussian-distributed emitter signals.

Though subspace methods provide only one paradigm for signal processing, their popularity and success in implementation may be attributed to the simple underlying model for the signal and noise. The assumptions on the model could range from the requirement of having a smaller number of signals than sensors to a noise field characterization which may be unrealistic for real data gathered from sensors. Nevertheless, the subspace methods provide a simple unified approach to tackle inherently difficult nonlinear parameter estimation problems, and their scopes continue to increase. For a recent review of approaches to tackle the nonlinear optimization problem which yield array localization signal processing techniques for estimating the number as well as parameters of narrowband and wideband sources, see [3]. For a recent review, in a unifying subspace-fitting framework of methodologies for analyzing a variety of subspace algorithms, see [4].

In this paper, we start with a model of second-order data (correlation matrix) generated from the samples of P received signals $\{x_i(n)\}$, $i = 1, 2, \ldots, P$. Each of these signals contains either a known or estimated number M of moderately correlated signals $y_k(n)$, $k = 1, 2, \ldots, M$, having distinct directions of arrivals θ_1, $\theta_2, \ldots, \theta_M$ with respect to a linear array of point sensors, embedded in additive independent and identically distributed Gaussian noise, as shown in (1) below. The assumption that the signal subspace dimension is either known or estimated is made in all subspace-based methods including MUSIC, ESPRIT, and GEESE. The dimension of the signal subspace is estimated by comparing the magnitudes of the eigenvalues of the correlation matrix and using a threshold. For $i = 1, 2, \ldots, P$, $j \triangleq \sqrt{-1}$, and $\omega_k \triangleq \pi \cos \theta_k$,

$$x_i(n) = \sum_{k=1}^{M} y_k(n)\exp[-j\omega_k(i-1)] + v_i(n), \quad n = 1, 2, \ldots, N. \quad (1)$$

* This research was supported by the Office of Naval Research under Grant N00014-92-J-1755.

1051-2004/94 $5.00

It is assumed that

$$0 < M \leq P - 1, \tag{2}$$

and noises at each of the P sensors are, ideally, zero-mean, wide-sense stationary uncorrelated random processes. Suppose that the common variance of the identically distributed uncorrelated noise processes is σ^2. Define

$$\underline{x}(n) = [x_1(n) \quad x_2(n) \cdots x_p(n)]^t \tag{3a}$$

and

$$\underline{y}(n) = [y_1(n) \quad y_2(n) \cdots y_M(n)]^t, \tag{3b}$$

where the superscript "t" denotes "transpose." The standard model for the $(P \times P)$ correlation matrix [1],

$$S = E[\underline{x}(n)\underline{x}^H(n)], \tag{4}$$

where the superscript "H" denotes "complex conjugate transpose," is

$$S = ARA^H + \sigma^2 I, \tag{5}$$

which is a $(P \times P)$ matrix, $R \triangleq E[\underline{y}(n)\underline{y}^H(n)]$ is an $(M \times M)$ nonsingular Hermitian matrix, and

$$A = \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j\omega_1} & \cdots & e^{-j\omega_M} \\ e^{-j(P-1)\omega_1} & \cdots & e^{-j(P-1)\omega_M} \end{bmatrix}. \tag{6}$$

The matrix I denotes throughout an identity matrix of *appropriate order;* to avoid notational clutter, we do not show the order explicitly in I. The above model provided in (1) with respect to the direction-of-arrival (DOA) problem applies also in several other situations, including the problem of parameter estimation of sinusoids.

2. MAIN RESULTS

2.1. Analysis of the Standard Model

The standard model for the $(P \times P)$ correlation matrix S was given in (5). From (4) and (5),

$$S = E[\underline{x}(n)\underline{x}^H(n)] = ARA^H + \sigma^2 I,$$

where A is shown in (6). On defining

$$\underline{a}(e^{-j\omega_i}) = [1 \quad e^{-j\omega_i} \cdots e^{-j(P-1)\omega_i}]^t, \tag{7}$$

the matrix A can be rewritten as

$$A = [\underline{a}(e^{-j\omega_1})\underline{a}(e^{-j\omega_2}) \cdots \underline{a}(e^{-j\omega_M})].$$

Since the ω_i's for $i = 1, 2, \ldots, M$ are distinct, the $(P \times M)$ matrix A is of full rank M, and because R is a nonsingular $(M \times M)$ Hermitian matrix, the matrix ARA^H must be a $(P \times P)$ nonnegative definite Hermitian matrix of rank M. Therefore, ARA^H has M nonzero real eigenvalues μ_i, $i = 1, 2, \ldots, M$. Since S is Hermitian and is modeled as in (5), it must be diagonalizable by a unitary matrix B [5]. Therefore,

$$B^H SB = B^H(ARA^H + \sigma^2 I)B$$
$$= \text{Diag}[(\mu_1 + \sigma^2) \cdots (\mu_M + \sigma^2)\sigma^2 \cdots \sigma^2] \tag{8}$$

is a $(P \times P)$ diagonal matrix. For notational brevity, we define below the $(M \times M)$ and $(P - M) \times (P - M)$ diagonal matrices, Δ_1 and Δ_2, respectively:

$$\Delta_1 \triangleq \text{Diag}[\lambda_1\lambda_2 \cdots \lambda_M], \quad \lambda_i = \mu_i + \sigma^2 \quad \text{for} \quad i = 1, \ldots, M \tag{9a}$$

$$\Delta_2 \triangleq \text{Diag}[\lambda_{M+1}\lambda_{M+2} \cdots \lambda_P], \quad \lambda_{M+i} = \sigma^2 \quad \text{for} \quad i = 1, \ldots, P - M. \tag{9b}$$

Denote the $(P \times M)$ matrix composed of the eigenvectors associated with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_M$ by B_S (signal subspace eigenvector matrix) and the $(P \times (P - M))$ matrix composed of the $(P - M)$ eigenvectors associated with the eigenvalues λ_i, $i = M + 1, M + 2, \ldots, P$ by B_N (noise subspace eigenvector matrix), so that

$$B = [B_S \; B_N]. \tag{10}$$

Some well-known results are rederived to facilitate the documentation of the novel results. From (10) and (8), since $B^H B = I$,

$$[B_S \; B_N]^H(ARA^H)[B_S \; B_N] + \sigma^2 I = \text{Diag}[\Delta_1 \; \Delta_2]. \tag{11}$$

The above equation leads to the following two equalities:

$$ARA^H B_S = B_S\Delta_1 - B_S(\sigma^2 I)$$
$$= B_S\text{Diag}[\mu_1 \cdots \mu_M] \tag{12a}$$

$$ARA^H B_N = B_N\Delta_2 - B_N(\sigma^2 I) = 0. \tag{12b}$$

Obviously, in (12a) and (12b), the I's denote identity matrices of orders M and $(P - M)$, respectively.

Since AR is of rank M, it follows from (12b) that the $(M \times (P-M))$ matrix $A^{H}B_{N}$ must satisfy

$$A^{H}B_{N} = 0. \tag{13}$$

Since $B_{S}^{H}B_{S} = I$, (12a) implies

$$(B_{S}^{H}A)R(A^{H}B_{S}) = \mathrm{Diag}[\mu_1, \mu_2, \ldots, \mu_M],$$

which in turn implies that the $(M \times M)$ matrix $A^{H}B_{S}$ is of nonsingular. Thus, it follows from (12a) that

$$A = B_{S}\mathrm{Diag}[\mu_1 \cdots \mu_M](A^{H}B_{S})^{-1}R^{-1} = B_{S}C, \tag{14}$$

where $C = \mathrm{Diag}[\mu_1 \cdots \mu_M](A^{H}B_{S})^{-1}R^{-1}$ is, obviously, nonsingular. It is seen later that the columns of C are related to the eigenvectors of the matrix whose eigenvalues estimate the parameters in the standard model of the signal correlation matrix.

Alternatives to (13) and (14) may also be obtained from (5). Since in practice the matrix B violates the assumptions of the data model, the use of a modified B (like the matrix obtained from B by complex conjugation of its elements) is likely to amplify the errors in parameter estimation. Therefore, the alternate derivation given next is necessary, where the counterparts of (13) and (14) are obtained in (20) and (21) below directly from the model. This type of derivation minimizes the ambiguity that is likely for the location of the roots of a polynomial to be on the unit circle when certain necessary conditions on the coefficients of the polynomial are not satisfied as explained later in Subsection 2.2 below. Let the exchange matrix be denoted by

$$J = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Premultiplying and postmultiplying both sides of (5) by J, we have

$$JSJ = E[\{J\underline{x}(n)\}\{J\underline{x}(n)\}^{H}] = (JA)R(JA)^{H} + \sigma^2 I. \tag{15}$$

JA is rearranged as

$$\begin{aligned} JA &= [J\underline{a}(e^{-j\omega_1}) \cdots J\underline{a}(e^{-j\omega_M})] \\ &= [e^{-j(P-1)\omega_1}\underline{a}(e^{j\omega_1}) \cdots e^{-j(P-1)\omega_M}\underline{a}(e^{j\omega_M})] \\ &= [\underline{a}(e^{j\omega_1}) \cdots \underline{a}(e^{j\omega_M})] \\ &\quad \times \mathrm{Diag}[e^{-j(P-1)\omega_1} \cdots e^{-j(P-1)\omega_M}] \\ &= A_{+}\Lambda, \end{aligned}$$

where

$$\underline{a}(e^{j\omega_i}) = [1\ e^{j\omega_i} \cdots e^{j(P-1)\omega_i}]^{t},$$

$$A_{+} = [\underline{a}(e^{j\omega_1}) \cdots \underline{a}(e^{j\omega_M})],$$

and

$$\Lambda = \mathrm{Diag}[e^{-j(P-1)\omega_1} \cdots e^{-j(P-1)\omega_M}].$$

Using the above equation, (15) is rewritten as

$$JSJ = A_{+}(\Lambda R \Lambda^{H})A_{+}^{H} + \sigma^2 I = A_{+}R_{+}A_{+}^{H} + \sigma^2 I, \tag{16}$$

where $R_{+} = \Lambda R \Lambda^{H}$ is, obviously, nonsingular.

The matrix JSJ is Hermitian and since

$$\begin{aligned} B^{H}SB &= (JB)^{H}(JSJ)(JB) \\ &= \mathrm{Diag}[\lambda_1\lambda_2 \cdots \lambda_M\lambda_{M+1} \cdots \lambda_p] \end{aligned} \tag{17}$$

with $(JB)^{H}(JB) = (JB)(JB)^{H} = I$, therefore JB is the eigenvector matrix associated with eigendecomposition of JSJ. Furthermore, since $JB = [JB_{S}JB_{N}] = [B_{SJ}B_{NJ}]$ where $B_{SJ} = JB_{S}$ and $B_{NJ} = JB_{N}$, the $(P \times M)$ eigenvector matrix B_{SJ} is associated with λ_i's, $i = 1, 2, \ldots, M$ and the $(P \times (P-M))$ eigenvector matrix B_{NJ} is associated with λ_i's, $i = M+1, \ldots, P$. It follows from (16) and (17) that

$$\begin{aligned} [B_{SJ}\ B_{NJ}]^{H}(JSJ)[B_{SJ}\ B_{NJ}] &= [B_{SJ}\ B_{NJ}]^{H} \\ &\quad \times (A_{+}R_{+}A_{+}^{H})[B_{SJ}\ B_{NJ}] + \sigma^2 I = \mathrm{diag}[\Delta_1\Delta_2], \end{aligned}$$

or

$$\begin{aligned} A_{+}R_{+}A_{+}^{H}B_{SJ} &= B_{SJ}\Delta_1 - B_{SJ}(\sigma^2 I) \\ &= B_{SJ}\mathrm{Diag}[\mu_1 \cdots \mu_M] \end{aligned} \tag{18}$$

and

$$B_{NJ}^{H}A_{+}R_{+}A_{+}^{H} = \Delta_2 B_{NJ}^{H} - (\sigma^2 I)B_{NJ}^{H} = 0. \tag{19}$$

Since $R_{+}A_{+}^{H}$ is of rank M, it follows from (19) that the $((P-M) \times M)$ matrix, $B_{NJ}^{H}A_{+}$ must satisfy

$$B_{NJ}^{H}A_{+} = 0 \quad \text{or} \quad A_{+}^{T}B_{NJ}^{*} = 0 \tag{20}$$

where the star superscript denotes the operation of complex conjugation. Since $B_{SJ}^{H}B_{SJ} = I$, (18) implies

$$(B_{SJ}^{H}A_{+})R_{+}(A_{+}^{H}B_{SJ}) = \mathrm{Diag}[\mu_1 \ldots \mu_M],$$

which in turn implies that the $(M \times M)$ matrix $A_+^H B_{SJ}$ is nonsingular.

It then follows from (18) that

$$A_+ = B_{SJ}\text{Diag}[\mu_1 \cdots \mu_M](A_+^H B_{SJ})^{-1} R_+^{-1} = B_{SJ} C_+, \quad (21)$$

where $C_+ = \text{Diag}[\mu_1 \cdots \mu_M](A_+^H B_{SJ})^{-1} R_+^{-1}$ is, obviously, nonsingular. The role of C_+ in (21) is analogous to the role of C in (14). Since $A_+^T = A^H$, (13) leads to

$$A_+^t B_N = 0 \quad \text{and} \quad A_+^t B_{NJ}^* = 0. \quad (22)$$

Using (14), (21), and (22), an algorithm for estimating $e^{j\omega_i}$'s, $i = 1, 2, \ldots, M$, is derived next.

2.2. *Derivation of the Algorithm for Estimating Parameters*

After defining,

$$\underline{a}(z) = [1\ z \cdots z^{(P-1)}]^t, \quad (23)$$

it is clear from (22) that $z_i = e^{j\omega_i}$, $i = 1, 2, \ldots, M$, are the M roots of the $(P-1)$th-degree polynomial equations

$$[a(z)]^T \underline{b} = b_0 + b_1 z + b_2 z^2 + \cdots + b_{P-1} z^{P-1} = 0, \quad (24)$$

where $\underline{b}^t = [b_0\, b_1 \cdots b_{P-1}]$ is any one of the columns of B_N or B_{NJ}^*. Therefore,

$$A_+^T \underline{b} = 0. \quad (25)$$

Each column of B_N and B_{NJ}^* is associated with an equation similar to (24). The set of these $2(P-M)$ equations can be written as

$$[a(z)]^t [\underline{b}_1 \cdots \underline{b}_{2(P-M)}] = 0,$$

and from (25),

$$A_+^t [\underline{b}_1 \cdots \underline{b}_{2(P-M)}] = 0. \quad (26)$$

Without loss of generality, it may be assumed that $b_{P-1} \neq 0$ in (24), for if $b_{P-1} = 0$, the monomial of highest degree is selected. Multiply (24) by z, divide throughout by b_{P-1}, and rearrange as follows:

$$z^P + \frac{b_{P-2}}{b_{P-1}} z^{P-1} + \cdots + \frac{b_1}{b_{P-1}} z^2 + \frac{b_0}{b_{P-1}} z = 0. \quad (27)$$

After associating a companion matrix D of order P with the polynomial on the left-hand side of (27),

$$D = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & -\frac{b_0}{b_{P-1}} & -\frac{b_1}{b_{P-1}} & \cdots & -\frac{b_{P-2}}{b_{P-1}} \end{bmatrix}, \quad (28)$$

it is clear that the polynomial equation in (27) may be rewritten as

$$\det[zI - D] = 0.$$

The desired roots, $z_i = e^{j\omega_i}$, $i = 1, 2, \ldots, M$, on the unit circle are also the M unit magnitude eigenvalues z_i, $i = 1, 2, \ldots, M$ of D. For each such eigenvalue, the associated eigenvector is

$$\underline{a}(z_i) = [1\ z_i\ z_i^2 \cdots z_i^{P-1}]^t. \quad (29)$$

Therefore, the following eigenequations hold.

$$D\underline{a}(z_i) = z_i \underline{a}(z_i), \quad i = 1, 2, \ldots, M. \quad (30)$$

The set of equations in (30) may be rewritten as

$$D[\underline{a}(z_1) \cdots \underline{a}(z_M)] = [\underline{a}(z_1) \cdots \underline{a}(z_M)]\Delta_z, \quad (31)$$

where

$$\Delta_z = \text{Diag}[z_1\ z_2 \cdots z_M]. \quad (32)$$

Since $z_i = e^{j\omega_i}$, $i = 1, 2, \ldots, M$,

$$A_+ = [\underline{a}(z_1)\ \underline{a}(z_2) \cdots \underline{a}(z_M)], \quad (33)$$

and, therefore, (31) is expressible as

$$DA_+ = A_+ \Delta_z, \quad (34)$$

where Δ_z is the diagonal matrix containing the M solutions desired. The theorem, stated and proved below, shows how information from the signal and noise subspace eigenvectors provides an estimate of the parameters. Before applying the theorem, it is necessary to construct from the standard model of the $(P \times P)$ correlation matrix S, the $(P \times M)$ eigenvector matrix B_S, whose columns span the signal subspace, and the $(P \times (P-M))$ eigenvector matrix B_N, whose columns span the noise subspace. Then the $(P \times P)$ companion matrix D shown in (28) is constructed

from B_N, while the $(M \times M)$ matrix $(JB_S)^H D(JB_S)$ is generated from B_S and D.

THEOREM 1. *The $(M \times M)$ matrix $(JB_S)^H D(JB_S)$ is nonsingular, whose eigenvalues uniquely estimate the M distinct parameters $e^{j\omega_i}$, $i = 1, 2, \ldots, M$, in the standard model of the signal correlation matrix. It also holds for $B_S^t D B_S^*$.*

Proof. The matrix D in (28) can be written as

$$D = \begin{bmatrix} 0 & I \\ 0 & \underline{g}^t \end{bmatrix},$$

where I is the identity matrix of order $(P-1)$ and $\underline{g}^t$ is a row vector having $(P-1)$ elements.

A_+ can be decomposed into

$$A_+ = \begin{bmatrix} A_1 \\ h_1 \end{bmatrix} = \begin{bmatrix} h_2 \\ A_2 \end{bmatrix},$$

where

$$h_1 = [e^{j(P-1)\omega_1} \cdots e^{j(P-1)\omega_M}] \quad \text{and} \quad h_2 = [1 \cdots 1]. \tag{35}$$

Using (21),

$$\begin{aligned} B_{SJ}^H D B_{SJ} &= (C_+^{-1})^H A_+^H D A_+ C_+^{-1} \\ &= (C_+^{-1})^H [A_1^H h_1^H] \begin{bmatrix} 0 & I \\ 0 & \underline{g}^t \end{bmatrix} \begin{bmatrix} h_2 \\ A_2 \end{bmatrix} C_+^{-1} \\ &= (C_+^{-1})^H [A_1^H A_2 + h_1^H \underline{g}^t A_2] C_+^{-1}. \end{aligned}$$

From (35) it follows that $A_2 = A_1 \Delta_e$ where

$$\Delta_e = \operatorname{diag}[e^{j\omega_1} \cdots e^{j\omega_M}]. \tag{36}$$

Equations (25), (27), and (28) imply that

$$[-\underline{g}^t \; 1] A_+ = [-\underline{g}^t \; 1] \begin{bmatrix} A_1 \\ h_1 \end{bmatrix} = 0 \quad \text{or} \quad \underline{g}^t A_1 = h_1.$$

Therefore,

$$\begin{aligned} B_{SJ}^H D B_{SJ} &= (C_+^{-1})^H [A_1^H A_1 \Delta_e + h_1^H h_1 \Delta_e] C_+^{-1} \\ &= (C_+^{-1})^H A_+^H A_+ \Delta_e C_+^{-1} \\ &= C_+ \Delta_e C_+^{-1}, \end{aligned} \tag{37}$$

since $(C_+^{-1})^H A_+^H A_+ = (C_+^{-1})^H (C_+^H B_{SJ}^H B_{SJ} C_+) = C_+$.

Thus, we see that $B_{SJ}^H D B_{SJ} = C_+ \Delta_e C_+^{-1}$ holds exactly, and since C_+ is nonsingular, the eigenvalues of $B_{SJ}^H D B_{SJ}$ uniquely correspond to the diagonal elements in Δ_e which are the desired parameters. It is straightforward to see from (14) and $A_+ = A^*$ that the $(M \times M)$ matrix $B_S^t D B_S^*$ generated by replacing B_{SJ} with B_S^* in (37) is also nonsingular, whose eigenvalues are $e^{j\omega_i}$, $i = 1, 2, \ldots, M$.

The eigenvectors of the $(M \times M)$ matrix $(JB_S)^H D(JB_S)$ which can be used in the estimation of R_+ or R may be related to the matrix C_+ in (21) in a manner stated and proved in Theorem 2 below.

THEOREM 2. *The $(M \times M)$ matrix E of the M eigenvectors of the matrix $(JB_S)^H D(JB_S)$, determined subject to the constraint,*

$$(JB_S)_1 E = [1 \quad 1 \cdots 1], \tag{38}$$

where $(JB_S)_1$ denotes the first row of (JB_S), is the matrix C_+ satisfying the relation, $A_+ = B_{SJ} C_+ = JB_S C_+$ in (21).

Proof. Let z_i, $i = 1, 2, \ldots, M$, represent the distinct eigenvalues of $(JB_S)^H D(JB_S)$. Since $(JB_S)^H D(JB_S)$ is nonsingular, therefore, the matrix $[(JB_S)^H D(JB_S) - z_i I]$ is of rank $M-1$, and nullity 1. Consequently, the eigenvector $\underline{e}_i$ in the eigenequation,

$$[(JB_S)^H D(JB_S) - z_i I] \underline{e}_i = 0, \tag{39}$$

has only one nonzero free variable u_i. Therefore, for $i = 1, 2, \ldots, M$

$$\underline{e}_i = (\underline{e}_i^f) u_i, \tag{40}$$

where $\underline{e}_i^f$ is an $(M \times 1)$ vector of constants. Consequently,

$$\begin{aligned} E &= [\underline{e}_1 \; \underline{e}_2 \cdots \underline{e}_M] \\ &= [\underline{e}_1^f \; \underline{e}_2^f \cdots \underline{e}_M^f] \operatorname{diag}(u_1 u_2 \cdots u_M). \end{aligned} \tag{41}$$

Since E is constrained as in (38), therefore following the substitution of (41) into (38) it follows, for $i = 1, 2, \ldots, M$ that

$$u_i = \frac{1}{[JB_S]_1 \underline{e}_i^f}. \tag{42}$$

Since the u_i's are unique, E is also uniquely determined. If $(JB_S)_i$ denotes the ith row of JB_S for $i = 1, 2, \ldots, P$ then from (21)

$$A_+ = \begin{bmatrix} (JB_S)_1 C_+ \\ \vdots \\ (JB_S)_P C_+ \end{bmatrix}. \tag{43}$$

Note that each element in the first row of A_+ must be 1. To satisfy this condition, replace C_+ by E in (43) and apply the constraint in (38) to get

$$A_+ = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ & (JB_S)_2 E & & \\ & \vdots & & \\ & (JB_S)_P E & & \end{bmatrix}. \qquad (44)$$

Clearly, then, E is the unique matrix C_+ satisfying (21).

Equation (26) provides $2(P-M)$ eigenequations through Theorem 1 to decide only one set of eigenvalues. There may be several ways for solving the overdetermined eigenvalue problem. In the example given below, simply the mean of the $2(P-M)$ coefficient sets in (26) was used as $\underline{b}$ to give one eigenequation.

3. ILLUSTRATIVE EXAMPLE AND COMPARISONS

The example discussed here is the one considered in [6], where the MUSIC algorithm was applied to estimate the parameters of a real sinusoidal signal embedded in zero-mean white Gaussian noise under various signal-to-noise ratios (SNRs). It becomes essential to show how this type of problem may be fitted to the model in (5) so that the technical devices of the previous section may be applied. Let the number of complex sinusoids be M. Then, the received signal at one sensor is

$$x_1(n) = \sum_{k=1}^{M} s_k(n) e^{j\omega_k n} + v_1(n), \qquad (45)$$

where each discrete sinusoid whose angular frequency ω_k is required to be estimated is represented by $s_k(n)e^{-j\omega_k n}$. The component $s_k(n)$, comprised of the magnitude and phase random variables, is stochastic. The magnitude and phase are characterized by probability distribution functions which permit the process $\{s_k(n)\}$ to be viewed as zero-mean wide-sense stationary. In (45), $v_1(n)$ is a random variable which is associated with a zero-mean wide-sense stationary uncorrelated random process having a variance σ^2. Equation (45) may be recast in the form

$$x_1(n) = [e^{-j\omega_1 n} \quad e^{-j\omega_2 n} \cdots e^{-j\omega_M n}] \times \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_M(n) \end{bmatrix} + v_1(n),$$

which for N consecutive received samples leads, immediately, to

$$\underline{x}_1(n) \triangleq [x_1(n) \quad x_1(n+1) \cdots x_1(n+N-1)]^t = A\underline{s}(n) + \underline{v}(n), \qquad (46)$$

where

$$\underline{s}(n) \triangleq [e^{-j\omega_1 n}s_1(n) \quad e^{-j\omega_2 n}s_2(n) \cdots e^{-j\omega_M n}s_M(n)]^t,$$

$$\underline{v}(n) = [v(n) \quad v(n+1) \cdots v(n+N-1)]^t.$$

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-j\omega_1} & e^{-j\omega_2} & \cdots & e^{-j\omega_N} \\ \vdots & & & \vdots \\ e^{-j(N-1)\omega_1} & e^{-j(N-1)\omega_2} & \cdots & e^{-j(N-1)\omega_N} \end{bmatrix}.$$

Therefore, subject to the assumptions made

$$E[\underline{x}_1(n)\underline{x}_1^H(n)] = E[(A\underline{s}(n) + \underline{v}(n)) \times (A\underline{s}(n) + \underline{v}(n))^H] = AE[\underline{s}(n)\underline{s}^H(n)]A^H + \sigma^2 I. \qquad (47)$$

The correlation matrix, $E[\underline{x}_1(n)\underline{x}_1^H(n)]$, on the left-hand side of (47) may be estimated by the $M \times M$ matrix

$$S = X^H X, \qquad (48)$$

where for the N samples of the received signal, $\{x_1(0) \quad x_1(1) \cdots x_1(N-1)\}$, the $(N-P+1) \times P$ matrix X is given by

$$X = \begin{bmatrix} x(P-1) & x(P-2) & \cdots & x(0) \\ x(P) & x(P-1) & \cdots & x(1) \\ \vdots & & & \\ x(N-1) & x(N-2) & \cdots & x(N-P) \end{bmatrix} \qquad (49)$$

To the problem solved in [6], we apply here not only the new method proposed through Theorem 1 but also another subspace-based method called GEESE. It is also pointed out that the estimates of the parameters in [6] are inferred from a graph. Here, we derived the estimates by applying the Root-MUSIC algorithm [7] in addition to the GEESE algorithm. The Root-MUSIC method requires the evaluation of the roots of the equation

$$\underline{a}^t(z) B_N B_N^H \underline{a}(z^{-1}) = 0, \qquad (50)$$

where

$$\underline{a}(z) = [1 \quad z \cdots z^{P-1}]^t. \qquad (51)$$

The GEESE method [1, pp. 47–76] requires the computation of the singular values of the matrix pencil $\{B_1, B_2\}$, with

$$B_1 = [B_{S1}^t \quad B_{S2}^t \cdots B_{S(P-1)}^t]^t \tag{52a}$$

$$B_2 = [B_{S2}^t \quad B_{S3}^t \cdots B_{SP}^t]^t, \tag{52b}$$

where B_{Si} denotes the ith row, $i = 1, 2, \ldots, P$, of the $(P \times M)$ signal subspace eigenvector matrix B_S.

The problem considered in [6] involves the received data $x(n)$, $n = 0, 1, \ldots, 6$, modeled by

$$x(n) = \frac{1}{\sqrt{2}} e^{j\omega_1 n} + \frac{1}{\sqrt{2}} e^{j\omega_2 n} + v(n).$$

The parameters ω_1 and ω_2 are to be estimated from the data matrix

$$X = \begin{bmatrix} x(3) & x(2) & x(1) & x(0) \\ x(4) & x(3) & x(2) & x(1) \\ x(5) & x(4) & x(3) & x(2) \\ x(6) & x(5) & x(4) & x(3) \end{bmatrix}. \tag{53}$$

The correlation matrix S is generated from X through the equation,

$$S = X^H X.$$

Three cases are considered. The data matrices are given for SNR values of 10, 20, and 30 dB. The estimates for ω_1 and ω_2 are obtained by applying the new algorithm presented here, followed by Root-MUSIC and GEESE algorithms. The correct values of the parameters chosen in the simulation were

$$\omega_1 = -1 \quad \text{and} \quad \omega_2 = +1. \tag{54}$$

Consequently, roots z_1 and z_2 in the z-plane are on the unit circle, $|z| = 1$, where $z_i = \exp(j\omega_i)$, $i = 1, 2$. Since the data is noisy, the estimated roots need not lie on a circle of unit radius. Therefore, the radii, r_1 and r_2, each of which ideally should be unity, as well as the angular frequencies are given for each of the three methods applied to solve the problem.

Case 1 (SNR = 10 dB). The data matrix X is

−0.99313	0.42526	1.82929	1.31736
−0.90984	−0.99313	0.42526	1.82929
−0.87901	−0.90984	−0.99313	0.42526
−0.21668	−0.87901	−0.90984	−0.99313.

The correlation matrix S is

2.63373	1.47147	−1.13353	−3.13129
1.47147	2.76762	2.05893	−0.77045
−1.13353	2.05893	5.34126	3.66901
−3.13129	−0.77045	3.66901	6.24889.

The eigenvector matrix associated with S is

−0.47835	0.68927	0.40616	−0.36210
0.71288	0.09170	0.69515	0.01253
−0.49272	−0.35650	0.54186	0.58010
0.14212	0.62402	−0.24122	0.72953.

The eigenvalues of S are

0.20357 0.58090 5.49965 10.70738.

The estimates of ω_1 and ω_2 as well as the values of r_1 and r_2, each of which, ideally, should be unity are given below for each of the three algorithms.

Algorithm	r_1	ω_1	r_2	ω_2
New	0.90905	−1.00776	0.90905	1.00776
Root-MUSIC	0.85560	−0.90655	0.85560	0.90655
GEESE	0.87124	−0.90967	0.87124	0.90967

Case 2 (SNR = 20 dB). The data matrix X is

−0.94613	0.40878	1.50696	1.14561
−1.24504	−0.94613	0.40878	1.50696
−0.68038	−1.24504	−0.94613	0.40878
0.45395	−0.68038	−1.24504	−0.94613.

The correlation matrix S is

3.11427	1.32945	−1.85619	−3.66774
1.32945	3.07530	2.25433	−0.82270
−1.85619	2.25433	4.88332	3.13361
−3.66774	−0.82270	3.13361	4.64561.

The eigenvector matrix associated with S is

−0.50255	0.62403	0.34020	−0.49224
0.65479	0.13263	0.74396	0.01380
−0.55649	−0.28544	0.53005	0.57261
0.09496	0.71521	−0.22324	0.65546.

The eigenvalues of S are

0.01975 0.04224 5.53627 10.12025.

The estimates of ω_1 and ω_2 as well as the values of r_1 and r_2 are given below for each of the three algorithms.

Algorithm	r_1	ω_1	r_2	ω_2
New	0.97229	−1.00502	0.97229	1.00502
Root-MUSIC	0.94576	−0.97266	0.94576	0.97266
GEESE	0.94879	−0.97364	0.94879	0.97364

Case 3 (SNR = 30 dB). The data matrix X is

$$\begin{matrix} -0.93126 & 0.40357 & 1.40503 & 1.09130 \\ -1.35104 & -0.93126 & 0.40359 & 1.40503 \\ -0.61757 & -1.35104 & -0.93126 & 0.40357 \\ 0.66602 & -0.61757 & -1.35104 & -0.93126. \end{matrix}$$

The correlation matrix S is

$$\begin{matrix} 3.51753 & 1.30539 & -2.17842 & -3.78401 \\ 1.30539 & 3.23682 & 2.28371 & -0.83815 \\ -2.17842 & 2.28371 & 4.82955 & 2.98271 \\ -3.78401 & -0.83815 & 2.98271 & 4.19516. \end{matrix}$$

The eigenvector matrix associated with S is

$$\begin{matrix} -0.50987 & 0.58999 & 0.32172 & -0.53707 \\ 0.63320 & 0.14285 & 0.76061 & 0.01142 \\ -0.57720 & -0.26445 & 0.52162 & 0.56992 \\ 0.07703 & 0.74938 & -0.21420 & 0.62178. \end{matrix}$$

The eigenvalues of S are

$$0.00196 \quad 0.00369 \quad 5.59116 \quad 10.18224.$$

The estimates of ω_1 and ω_2 as well as the values of r_1 and r_2 are given below for each of the three algorithms.

Algorithm	r_1	ω_1	r_2	ω_2
New	0.99178	−1.00201	0.99178	1.00201
Root-MUSIC	0.98199	−0.99149	0.98199	0.99149
GEESE	0.98288	−0.99182	0.98288	0.99182

4. CONCLUSIONS

The MUSIC and Root-MUSIC-based algorithms explicitly use the eigenvectors spanning the noise subspace to estimate the direction-of-arrivals and the angular frequencies of sinusoids. Other subspace methods such as GEESE and ESPRIT use the signal subspace eigenvectors in the parameter estimation problems mentioned above. The present method uses both the noise and signal subspace eigenvectors for the stated purpose as seen from Theorem 1. Specifically, the matrix D is computed from the noise subspace eigenvectors while the matrix $(JB_S)^H D(JB_S)$, whose eigenvalues have to be computed, also depends on the signal subspace eigenvectors explicitly.

The new method is seen to give better estimates for ω_1 and ω_2 than Root-MUSIC and GEESE for the benchmark problem discussed in this paper. It is reasonable to expect that the statistics of the data can influence the accuracy of the results derived from implementation of the various algorithms, especially when the SNR is low. However, the presented method contributes to the popular subspace-theory-based techniques through the development of a new algorithm which works better in some situations. The conditions on the data under which such improved performance may be expected and optimal solution schemes for the overdetermined eigenvalue problem are currently under study. The scopes for generalizing the approach to the parameter estimation problems for two-dimensional sinusoids in noise is also worth investigating.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the constructive comments by Dr. Dennis Ricker and Dr. Stephan Schell.

REFERENCES

1. Pillai, S. U. *Array Signal Processing.* Springer-Verlag, New York, 1989.
2. Ottersten, B., Viberg, M., and Kailath, T. Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data. *IEEE Trans. Signal Process.* **40** (Mar. 1992), 590–600.
3. Hurt, N. E. Maximum likelihood estimation and MUSIC in array localization signal processing: A review. *Multidimensional Systems Signal Process.* **1** (1990), 279–325.
4. Paulraj, A., Ottersten, B., Roy, R., Swindlehurst, A., Xu, G., and Kailath, T. Subspace methods for directions-of-arrival estimation. In *Signal Processing and Its Applications* (N. K. Bose and C. R. Rao, Eds.), Chap. 16; Handbook of Statistics. North-Holland, Amsterdam, 1993.
5. Rao, C. R., and Mitra, S. K. *Generalized Inverse of Matrices and Its Applications.* Wiley, New York, 1971.
6. Haykin, S. Eigendecomposition-based high-resolution algorithms, *Modern Filters,* Chap. 13. MacMillan, New York, 1989.
7. Rao, Bhaskar D., and Hari, K. V. S. Performance analysis of Root-MUSIC. *IEEE Trans. Acoust. Speech Signal Process.* **17** (Dec. 1989), 1939–1949.

JONGTAE CHUN was born on June 13, 1962. He received his B.S. and M.S. in instrumentation and control at Seoul National University in 1983 and 1985, respectively. Between 1985 and 1987, he worked as a researcher at Korea Data Communication Corpora-

tion, Seoul, and from 1987 to 1991, he was a full-time consultant at Korea Productivity Center, Seoul. Since September 1991, he has been a graduate student in electrical engineering at The Pennsylvania State University, University Park, where he is a research assistant in the Spatial and Temporal Signal Processing Center, working toward his Ph.D. degree. His current research interests are in signal and image processing.

N. K. BOSE was born August 19, 1940. He received the B Tech (Hons.) degree, the M.S. degree, and the Ph.D. degree from the Indian Institute of Technology, Kharagpur, Cornell University, Ithaca, NY, and Syracuse University, Syracuse, NY, respectively. He is currently the HRB-Systems Professor of Electrical Engineering and the Director of The Spatial and Temporal Signal Processing Center at The Pennsylvania State University, University Park. Dr. Bose is the author of *Applied Multidimensional Systems Theory* (Van Nostrand, 1982), *Digital Filters* (North-Holland, 1985; Krieger Publishing, 1993), and the editor as well as the main author of *Multidimensional Systems: Progress, Directions and Open Problems* (Reidel, 1985). He was the Guest Editor of the *Proceedings of the IEEE* Special Issue on Multidimensional Systems in June 1977 and the editor of the IEEE Press Selected Reprint Series book on *Multidimensional Systems: Theory and Applications* (1979). He also served as Guest Editor of a Special Issue on *Aspects of Spatial and Temporal Signal Processing* published by Circuits, Systems, and Signal Processing (Birkhaeuser Boston, 1984). Dr. Bose was a Guest Co-Editor for a Special Issue on *Linear Algebra in Electrical Engineering* which was published by Linear Algebra and Its Applications in January 1988. Dr. Bose is a co-editor of the book on *Signal Processing and Its Applications*, published by North-Holland in 1993. He has been the Editor-in-Chief of the *Journal on Multidimensional Systems and Signal Processing* since 1990. He has been a Fellow of IEEE since 1981. His current research interests are in multidimensional signal processing, multitarget tracking, and neural networks.

Reprinted from Digital Signal Processing 4, 40–48 (1994)

DIGITAL SIGNAL PROCESSING 5, 58–76 (1995)

Parameter Estimation via Signal-Selectivity of Signal-Subspace (PESS) and Its Application to 2-D Wavenumber Estimation*

Jongtae Chun and N. K. Bose†

Department of Electrical Engineering, The Spatial and Temporal Signal Processing Center, The Pennsylvania State University, University Park, Pennsylvania 16802

Chun, J., and Bose, N. K., Parameter Estimation via Signal-Selectivity of Signal-Subspace (PESS) and Its Application to 2-D Wavenumber Estimation, *Digital Signal Processing* 5 (1995), 58–76.

In this paper, three new methods called the PESS methods are presented. The first is a 1-D based 2-D PESS method for estimating 2-D wavenumbers. This provides the basic framework for all the PESS methods. The second is the Pairing-PESS method which offers efficient pairing in all 1-D based 2-D methods. The third is the Direct 2-D PESS method which, though, in the category of direct 2-D methods in the sense that it is not based on any 1-D signal processing technique, is, nevertheless, computationally efficient even when compared with the previous 1-D based 2-D algorithms. This algorithm is different from the other direct 2-D algorithm in that it does not depend on spectral search in 2-D parameter space. Examples are presented to illustrate the methods and a comparison is made with the matrix enhancement matrix pencil method recently proposed.

1. INTRODUCTION

Extensions of 1-D subspace methods with the objective of estimating the wavenumber of a 2-D sinusoidal signal in noise as well as the direction-of-arrival (DOA), have been studied. The obstacle to finding computationally efficient high-resolution algorithms in such cases is the factorability problem associated with bivariate polynomials in ROOT-MUSIC and ROOT-MIN-NORM-based methods (the zeros of bivariate polynomials trace continuous algebraic curves [1]; see [2] for the testing for zeros on the distinguished boundary of an unit polydisc of a multivariate polynomial) and the so-called problem of pairing bivariate polynomial zeros in 2-D generalizations of ESPRIT. The 2-D parameter estimation problem referred to above has been tackled by two approaches. The first approach (*1-D based 2-D approach*) requires the decomposition of a 2-D problem into two separate 1-D problems followed by the construction of a set of 2-D solutions from the set of 1-D solutions by pairing algorithms. The second method (*direct 2-D approach*) solves the 2-D problem directly by searching for spectral peaks or manipulating optimization algorithms in multidimensional parameter space. Although the former is attractive computationally, it results in suboptimal solutions and suffers from the difficulty in implementing pairing. The latter can be modified to produce optimal solutions but the computation cost is very high.

A computationally efficient 1-D based 2-D wavenumber estimation method was proposed in [5]. This method identifies the state equation form inherent in the 2-D covariance data matrix. The given sample data matrix is singular value decomposition-decomposed (SVD-decomposed) to yield the factors of the model. This model gives the two state transition matrices in diagonal form. The diagonal elements of each are the desired 1-D signal zeros in each of the two dimensions. This method, called the state-space method, assumes distinct frequencies in either dimension and ignores noise. The subsequent pairing procedure uses the amplitudes of the signals. A method similar to the state-space method, called the matrix approximation method, was proposed in [6], where the original covariance data matrix is reconstructed in a least squares sense.

* This research was supported by the Office of Naval Research under Grant N00014-92-J-1755.

† E-mail: nkb@stspnkb.psu.edu.

1051-2004/95 $5.00

Separable extensions of ESPRIT to the 2-D case have been discussed as 1-D based 2-D DOA estimation methods. Procrustes rotations (PRO)-based ESPRIT, proposed for compensating imperfect array data, was applied separately and the estimated azimuth and elevation angles were paired by evaluating the array response vectors corresponding to all the possible combinations of the estimates and then searching for those which are the closest to the signal subspace [7]. In [16], an implementation of ESPRIT was given based on delta array geometry, where phase differences are estimated and used as the criteria for pairing the parameters computed separately. Assuming an array of sensor triplets with arbitrary displacements, two matrix pencil equations were set up and two sets of 1-D DOAs were computed in [15] using the SVD of three data matrices. Observing that pairing is automatic in the noise-free case, noise cancelation was tried through the estimation of the perturbation matrices for the noisy case. An extension of ESPRIT to 2-D case using the concept of marked subspace was proposed in [10]. This algorithm introduced a mark operator which assigns different amplitudes to the roots in each 1-D solution set but the same amplitude to each correct pair of roots. The paired solutions are identified by finding two roots associated with the same amplitude.

Prony's methods for estimating 2-D sinusoidal frequencies, amplitudes, and phases were proposed [4,17]. Particularly in [4], the signal model is identified to be expressible as a recognizable rational function [1, p. 269]. The coefficient fitting is performed from the best reduced rank least-squares approximate obtained via SVD. In [8] a 2-D wavenumber estimation technique based on a matrix pencil (MEMP) was proposed. The pairing procedure is similar to that in [7].

The Fourier method, the maximum-likelihood method, the maximum-entropy method, the autoregressive method and the MUSIC method for estimating wavenumbers or DOA, based on multidimensional search, were presented in [11–13]. 2-D Toeplitz Approximation Method (TAM) followed by 2-D MUSIC was studied in [14].

An extension of subspace-fitting method to the 2-D case was investigated in [9], where optimal and suboptimal procedures for estimating 2-D DOA were formulated. It was shown that for uncorrelated and correlated sources these methods were significantly better than MUSIC. The computation, however, can be prohibitively expensive without good initial guess of the estimates.

In this paper, three new methods called the PESS methods are presented. The first is a 1-D based 2-D PESS method for estimating 2-D wavenumbers. This provides the basic framework for all the PESS methods. The second is the Pairing-PESS method which offers efficient pairing in all 1-D based 2-D methods. The third is the Direct 2-D PESS method which, though, in the category of direct 2-D methods in the sense that it is not based on any 1-D signal processing technique, is, nevertheless, computationally efficient even when compared with the previous 1-D based 2-D algorithms. This algorithm is different from the other direct 2-D algorithms in that it does not depend on spectral search in 2-D parameter space.

2. PROBLEM FORMULATION

It is assumed that the 2-D array of discrete-valued data, $x(n_1, n_2)$, can be modeled as

$$x(n_1, n_2) = \sum_{i=1}^{M} a_i e^{j(\omega_{1i} \cdot n_1 + \omega_{2i} \cdot n_2)} + v(n_1, n_2). \quad (2.1)$$

The noise process $v(n_1, n_2)$ is assumed to be zero-mean and white over n_1 and n_2 with variance σ^2. The coefficient a_i represents the amplitude and phase of the ith 2-D sinusoid. The number M of 2-D sinusoids may be estimated by thresholding a set of eigenvalues of a covariance matrix as mentioned later. In this paper, it is assumed that M is known.

Assume that the two-tuple, $(\omega_{1i}, \omega_{2i})$, which represents the i-th wavenumber, is distinct for each i in $1 \leq i \leq M$, that is,

$$\begin{pmatrix} \omega_{1i} = \omega_{1k} \\ \omega_{2i} \neq \omega_{2k} \end{pmatrix}, \quad \begin{pmatrix} \omega_{1i} \neq \omega_{1k} \\ \omega_{2i} = \omega_{2k} \end{pmatrix},$$

$$\text{or} \quad \begin{pmatrix} \omega_{1i} \neq \omega_{1k} \\ \omega_{2i} \neq \omega_{2k} \end{pmatrix}, \quad \text{whenever } i \neq k.$$

Note that either ω_{1i} or ω_{2i} can be zero for any i. Without loss of generality, the trivial case when both are zero is excluded. Also the alias-free constraint is imposed, that is, ω_{mi} is unique over $(-\pi, \pi]$ for $m = 1, 2$ and $1 \leq i \leq M$.

The problem to be solved is formulated as follows: Given the $N_1 \times N_2$ data matrix

$$X = \begin{bmatrix} x(0,0) & x(0,1) & \cdots & x(0, N_2 - 1) \\ x(1,0) & x(1,1) & \cdots & x(1, N_2 - 1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N_1 - 1, 0) & x(N_1 - 1, 1) & \cdots & x(N_1 - 1, N_2 - 1) \end{bmatrix}, \quad (2.2)$$

estimate the M 2-D wavenumbers, $(\omega_{1i}, \omega_{2i})$, for $1 \leq i \leq M$, in paired form.

3. DERIVATION OF SUBSPACES FOR 2-D SINUSOIDAL SIGNALS

3.1. Correlation Matrix

The measurement model of Eq. (2.1) can be expressed, using the vector notation, as

$$x(n_1, n_2) = [e^{j(\omega_{11} \cdot n_1 + \omega_{21} \cdot n_2)}, \cdots, e^{j(\omega_{1M} \cdot n_1 + \omega_{2M} \cdot n_2)}] \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix} + v(n_1, n_2). \quad (3.1)$$

For selected integers P_1 and P_2 such that $M < P_1 < N_1$ and $M < P_2 < N_2$, the data $x(n_1 + k_1, n_2 + k_2)$, for $0 \leq k_1 \leq P_1 - 1$ and $0 \leq k_2 \leq P_2 - 1$, may be stacked as a $(P_1 \cdot P_2) \times 1$ column vector,

$$\mathbf{x}(n_1, n_2) = \begin{matrix} [x(n_1, n_2) & x(n_1 + 1, n_2) & \cdots & x(n_1 + P_1 - 1, n_2) \\ x(n_1, n_2 + 1) & x(n_1 + 1, n_2 + 1) & \cdots & x(n_1 + P_1 - 1, n_2 + 1) \\ \cdots & \cdots & \cdots & \cdots \\ x(n_1, n_2 + P_2 - 1) & x(n_1 + 1, n_2 + P_2 - 1) & \cdots & x(n_1 + P_1 - 1, n_2 + P_2 - 1)]^t, \end{matrix} \quad (3.2)$$

where the superscript t denotes transpose. Substituting the measurement model in Eq. (3.1) for each element of $\mathbf{x}(n_1, n_2)$, described in Eq. (3.2), we get

$$\mathbf{x}(n_1, n_2) = \begin{bmatrix} e^{j(\omega_{11}n_1 + \omega_{21}n_2)} & \cdots & e^{j(\omega_{1M}n_1 + \omega_{2M}n_2)} \\ e^{j(\omega_{11}(n_1+1) + \omega_{21}n_2)} & \cdots & e^{j(\omega_{1M}(n_1+1) + \omega_{2M}n_2)} \\ \vdots & \vdots & \vdots \\ e^{j(\omega_{11}(n_1+P_1-1) + \omega_{21}n_2)} & \cdots & e^{j(\omega_{1M}(n_1+P_1-1) + \omega_{2M}n_2)} \\ e^{j(\omega_{11}n_1 + \omega_{21}(n_2+1))} & \cdots & e^{j(\omega_{1M}n_1 + \omega_{2M}(n_2+1))} \\ e^{j(\omega_{11}(n_1+1) + \omega_{21}(n_2+1))} & \cdots & e^{j(\omega_{1M}(n_1+1) + \omega_{2M}(n_2+1))} \\ \vdots & \vdots & \vdots \\ e^{j(\omega_{11}(n_1+P_1-1) + \omega_{21}(n_2+1))} & \cdots & e^{j(\omega_{1M}(n_1+P_1-1) + \omega_{2M}(n_2+1))} \\ \vdots & \vdots & \vdots \\ e^{j(\omega_{11}n_1 + \omega_{21}(n_2+P_2-1))} & \cdots & e^{j(\omega_{1M}n_1 + \omega_{2M}(n_2+P_2-1))} \\ e^{j(\omega_{11}(n_1+1) + \omega_{21}(n_2+P_2-1))} & \cdots & e^{j(\omega_{1M}(n_1+1) + \omega_{2M}(n_2+P_2-1))} \\ \vdots & \vdots & \vdots \\ e^{j(\omega_{11}(n_1+P_1-1) + \omega_{21}(n_2+P_2-1))} & \cdots & e^{j(\omega_{1M}(n_1+P_1-1) + \omega_{2M}(n_2+P_2-1))} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}$$

$$+ \begin{matrix} [v(n_1, n_2) & v(n_1 + 1, n_2) & \cdots & v(n_1 + P_1 - 1, n_2) \\ v(n_1, n_2 + 1) & v(n_1 + 1, n_2 + 1) & \cdots & v(n_1 + P_1 - 1, n_2 + 1) \\ \cdots & \cdots & \cdots & \cdots \\ v(n_1, n_2 + P_2 - 1) & v(n_1 + 1, n_2 + P_2 - 1) & \cdots & v(n_1 + P_1 - 1, n_2 + P_2 - 1)]^t. \end{matrix} \quad (3.3)$$

Denoting the rightmost vector in Eq. (3.3) by $\mathbf{v}(n_1, n_2)$, the equation may be rewritten as

$$\mathbf{x}(n_1, n_2) = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j\omega_{11}} & \cdots & e^{j\omega_{1M}} \\ \vdots & \vdots & \vdots \\ e^{j\omega_{11}(P_1-1)} & \cdots & e^{j\omega_{1M}(P_1-1)} \\ e^{j\omega_{21}} & \cdots & e^{j\omega_{2M}} \\ e^{j\omega_{21}}e^{j\omega_{11}} & \cdots & e^{j\omega_{2M}}e^{j\omega_{1M}} \\ \vdots & \vdots & \vdots \\ e^{j\omega_{21}}e^{j\omega_{11}(P_1-1)} & \cdots & e^{\omega_{2M}}e^{j\omega_{1M}(P_1-1)} \\ \vdots & \vdots & \vdots \\ e^{j\omega_{21}(P_2-1)} & \cdots & e^{j\omega_{2M}(P_2-1)} \\ e^{j\omega_{21}(P_2-1)}e^{j\omega_{11}} & \cdots & e^{j\omega_{2M}(P_2-1)}e^{j\omega_{1M}} \\ \vdots & \vdots & \vdots \\ e^{j\omega_{21}(P_2-1)}e^{j\omega_{11}(P_1-1)} & \cdots & e^{j\omega_{2M}(P_2-1)}e^{j\omega_{1M}(P_1-1)} \end{bmatrix} \cdot \begin{bmatrix} a_1 e^{j(\omega_{11}n_1 + \omega_{21}n_2)} \\ a_2 e^{j(\omega_{12}n_1 + \omega_{22}n_2)} \\ \vdots \\ a_M e^{j(\omega_M n_1 + \omega_{2M}n_2)} \end{bmatrix} + \mathbf{v}(n_1, n_2). \quad (3.4)$$

After introducing notations,

$$\mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) = \begin{matrix} [1 & e^{j\omega_{1i}} \\ e^{j\omega_{2i}} & e^{j\omega_{2i}}e^{j\omega_{1i}} \\ \cdots & \cdots \\ e^{j\omega_{2i}(P_2-1)} & e^{j\omega_{2i}(P_2-1)}e^{j\omega_{1i}} \end{matrix} \begin{matrix} \cdots & e^{j\omega_{1i}(P_1-1)} \\ \cdots & e^{j\omega_{2i}}e^{j\omega_{1i}(P_1-1)} \\ \cdots & \cdots \\ \cdots & e^{j\omega_{2i}(P_2-1)}e^{j\omega_{1i}(P_1-1)}]^t \end{matrix} \tag{3.5}$$

and

$$A = [\mathbf{a}(e^{j\omega_{11}}, e^{j\omega_{21}})\mathbf{a}(e^{j\omega_{12}}, e^{j\omega_{22}}) \cdots \mathbf{a}(e^{j\omega_{1M}}, e^{j\omega_{2M}})], \tag{3.6}$$

Eq. (3.4) is simply written as

$$\mathbf{x}(n_1, n_2) = \mathbf{A} \cdot \mathbf{s}(n_1, n_2) + \mathbf{v}(n_1, n_2), \tag{3.7}$$

where $\mathbf{s}(n_1, n_2)$ is easily identified by comparing Eq. (3.4) and (3.7).

We see that $\mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}})$ in Eq. (3.5) may be expressed as,

$$\mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) = \mathbf{a}_2(e^{j\omega_{2i}}) \otimes \mathbf{a}_1(e^{j\omega_{1i}}), \tag{3.8}$$

where $\otimes$ denotes the Kronecker product (let $A = B \otimes C$, where B is of order $p \times q$ and C is of order $s \times t$; then, the order of A is $(ps) \times (qt)$ and $(A)_{s(i-1)+m),(t(j-1)+n)} = (B)_{i,j}(C)_{m,n}$) and

$$\begin{aligned} \mathbf{a}_1(e^{j\omega_{1i}}) &= [1\ e^{j\omega_{1i}} \cdots e^{j\omega_{1i}(P_1-1)}]^t, \\ \mathbf{a}_2(e^{j\omega_{2i}}) &= [1\ e^{j\omega_{2i}} \cdots e^{j\omega_{2i}(P_2-1)}]^t. \end{aligned} \tag{3.9}$$

Denoting the (i, k)-th element of the $M \times M$ matrix $\mathbf{s}(n_1, n_2)\mathbf{s}^H(n_1, n_2)$ by $s_{ik}(n_1, n_2)$, we have

$$s_{ik}(n_1, n_2) = a_i a_k e^{j(\omega_{1i}-\omega_{1k})n_1} e^{j(\omega_{2i}-\omega_{2k})n_2}. \tag{3.10}$$

If $\omega_{mi} \neq \omega_{mk}$, for $m = 1$ or 2, we see that

$$\begin{aligned} \left| \lim_{N_m \to \infty} \frac{1}{N_m} \sum_{n_m=0}^{N_m-1} e^{j(\omega_{mi}-\omega_{mk})n_m} \right| &\leq \lim_{N_m \to \infty} \frac{1}{N_m} \left| \sum_{n_m=0}^{N_m-1} e^{j(\omega_{mi}-\omega_{mk})n_m} \right| \\ &\leq \lim_{N_m \to \infty} \frac{1}{N_m} \left| \frac{1 - e^{j(\omega_{mi}-\omega_{mk})N_m}}{1 - e^{j(\omega_{mi}-\omega_{mk})}} \right| \\ &\leq \lim_{N_m \to \infty} \frac{1}{N_m} \frac{2}{|1 - e^{j(\omega_{2i}-\omega_{2k})}|} = 0. \end{aligned}$$

If $\omega_{mi} = \omega_{mk}$, for $m = 1$ or 2, we see that

$$\lim_{N_m \to \infty} \frac{1}{N_m} \sum_{n_m=0}^{N_m-1} e^{j(\omega_{mi}-\omega_{mk})n_m} = \lim_{N_m \to \infty} \frac{1}{N_m} \sum_{n_m=0}^{N_m-1} 1 = 1.$$

Since $(\omega_{1i}, \omega_{2i}) \neq (\omega_{1k}, \omega_{2k})$ whenever $i \neq k$, it follows that

$$\lim_{N_2 \to \infty} \lim_{N_1 \to \infty} \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} s_{ik}(n_1, n_2) = \delta_{(i-k)},$$

where δ is the Kronecker delta function. So, we may view the vector $\mathbf{s}(n_1, n_2)$ to consist of, as elements, uncorrelated random variables. Since the data has finite support, we assume that the source correlation matrix S defined by

$$S \triangleq E[\mathbf{s}(n_1, n_2)\mathbf{s}^H(n_1, n_2)], \tag{3.11}$$

is nonsingular, where the superscript H denotes complex-conjugate transpose. This assumption holds when we have moderate amounts of data, and the 2-D wavenumbers are not too close. The correlation matrix R of $\mathbf{x}(n_1, n_2)$ has the structure

$$R = E[\mathbf{x}(n_1, n_2)\mathbf{x}^H(n_1, n_2)] = ASA^H + \sigma^2 I, \tag{3.12}$$

where A and S are defined in Eqs. (3.6) and (3.11), respectively.

3.2. Estimation of the Correlation Matrix R

Construct a $P_1 \times P_2$ matrix $X(n_1, n_2)$ expressed by

$$X(n_1, n_2) = \begin{bmatrix} x(n_1, n_2) & x(n_1, n_2+1) & \cdots & x(n_1, n_2+P_2-1) \\ x(n_1+1, n_2) & x(n_1+1, n_2+1) & \cdots & x(n_1+1, n_2+P_2-1) \\ \vdots & \vdots & & \vdots \\ x(n_1+P_1-1, n_2) & x(n_1+P_1-1, n_2+1) & \cdots & x(n_1+P_1-1, n_2+P_2-1) \end{bmatrix}. \tag{3.13}$$

We know that, from Eq. (3.2) and (3.13),

$$\mathbf{x}(n_1, n_2) = \text{vec } X(n_1, n_2), \tag{3.14}$$

where the vec operator stacks the columns of a matrix into a vector. (If $H = [\mathbf{h}_1\mathbf{h}_2 \cdots \mathbf{h}_M]$, then vec $H = [\mathbf{h}_1^t\mathbf{h}_2^t \cdots \mathbf{h}_M^t]^t$.) Actually, $x(n_1, n_2)$ is a window through which we look at the data matrix X of Eq. (2.2). The window of size $P_1 \times P_2$ moves horizontally with indexing n_2 and vertically with indexing n_1 within the

data matrix X. Each windowed submatrix $X(n_1, n_2)$ of X is mapped to $\mathbf{x}(n_1, n_2)$ by Eq. (3.14). In fact, vec $X(n_1, n_2)$, for $0 \leq n_1 \leq N_1 - P_1$ and $0 \leq n_2 \leq N_2 - P_2$, is interpreted to be the realization of the random vector $\mathbf{x}(n_1, n_2)$. Define the reorganized data matrix X_r to be

$$X_r = \begin{bmatrix} \text{vec } X(0,0) & \text{vec } X(1,0) & \cdots & \text{vec } X(N_1-P_1,0) \\ \text{vec } X(0,1) & \text{vec } X(1,1) & \cdots & \text{vec } X(N_1-P_1,1) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \text{vec } X(0,N_2-P_2) & \text{vec } X(1,N_2-P_2) & \cdots & \text{vec } X(N_1-P_1,N_2-P_2) \end{bmatrix}, \quad (3.15)$$

obtained by stacking vec $X(n_1, n_2)$, for $0 \leq n_1 \leq N_1 - P_1$ and $0 \leq n_2 \leq N_2 - P_2$, as columns. Clearly X_r has P_1P_2 rows and $[(N_1 - P_1 + 1)(N_2 - P_2 + 1)]$ columns. Let J be the permutation matrix of appropriate order defined below.

$$J = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \iddots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}.$$

By analyzing $J(\mathbf{x}(n_1, n_2))^*$ similarly to the case of $\mathbf{x}(n_1, n_2)$, discussed in the previous subsection, we can show that JX_r^* can be considered to be a realization of $\mathbf{x}(n_1, n_2)$. Assuming that $\mathbf{x}(n_1, n_2)$ is correlation ergodic, the estimate $\hat{R}$ of the correlation matrix R is computed from

$$\hat{R} = \frac{1}{2(N_1 - P_1 + 1)(N_2 - P_2 + 1)}(X_r X_r^H + JX_r^* X_r^t J). \quad (3.16)$$

From Eqs. (3.12) and (3.16), since

$$E[\hat{R}] = R,$$

the estimator in Eq. (3.16) is unbiased.

3.3. Derivation of Subspaces

The $(P_1P_2) \times M$ matrix A in Eq. (3.6) has columns, $\mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) \triangleq \mathbf{a}_2(e^{j\omega_{2i}}) \otimes \mathbf{a}_1(e^{j\omega_{1i}})$, for $1 \leq i \leq M$, each of which can be expressed by

$$\mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) = \begin{bmatrix} \mathbf{a}_1(e^{j\omega_{1i}}) \\ e^{j\omega_{2i}}\mathbf{a}_1(e^{j\omega_{1i}}) \\ \vdots \\ e^{j\omega_{2i}(P_2-1)}\mathbf{a}_1(e^{j\omega_{1i}}) \end{bmatrix}. \quad (3.17)$$

Since the first P_1 rows of A are given by $[\mathbf{a}_1(e^{j\omega_{11}}) \cdots \mathbf{a}_1(e^{j\omega_{1M}})]$, and $P_1 > M$, A is of full rank, provided the ω_{1i}'s, for $1 \leq i \leq M$, are distinct. When the ω_{1i}'s are not distinct, consider the linear equation

$$\sum_{i=1}^{M} c_i \mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) = 0$$

which, after using Eq. (3.8), becomes

$$\sum_{i=1}^{M} c_i \mathbf{a}_2(e^{j\omega_{2i}}) \otimes \mathbf{a}_1(e^{j\omega_{1i}}) = 0.$$

Suppose that all the column vectors of A in Eq. (3.6), whose first element $e^{j\omega_{1i}}$ is the same as $e^{j\omega_{11}}$, are $\mathbf{a}(e^{j\omega_{11}}, e^{j\omega_{2i_k}})$, for $1 \leq k \leq m$. Then the above equation is reorganized as

$$[\sum_{k=1}^{m} c_{i_k} \mathbf{a}_2(e^{j\omega_{2i_k}})] \otimes \mathbf{a}_1(e^{j\omega_{11}}) + \sum_{i \in \mathcal{F}_1} c_i \mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) = 0,$$

where $\omega_{1i} \neq \omega_{11}$ for any $i \in \mathcal{F}_1$. Since $\mathbf{a}_1(e^{j\omega_{11}})$ and $\mathbf{a}_1(e^{j\omega_{1i}})$ are linearly independent for $i \in \mathcal{F}_1$, and $\mathbf{a}_1(\cdot)$ appears as in Eq. (3.17), it must follow that $\sum_{k=1}^{m} c_{i_k} \mathbf{a}_2(e^{j\omega_{2i_k}}) = 0$, which in turn implies that $c_{i_k} = 0$ for $1 \leq k \leq m$, due to the fact that the ω_{2i_k}'s are distinct from the assumption that $(\omega_{1i}, \omega_{2i}) \neq (\omega_{1k}, \omega_{2k})$ whenever $i \neq k$. Continuing the above argument for each $e^{j\omega_{1i}}$, we get $c_i = 0$ for $1 \leq i \leq M$. Thus, all the columns of A are linearly independent and so A is of full rank, if $(\omega_{1i}, \omega_{2i}) \neq (\omega_{1k}, \omega_{2k})$ whenever $i \neq k$, which is consistent with the results of [4].

Since the correlation matrix R is Hermitian, and ASA^H has rank M, the eigendecomposition of R has the structure described by [19, p. 168]

$$B^H R B = B^H A S A^H B + \sigma^2 I$$
$$= \text{Diag}[\mu_1 + \sigma^2 \cdots \mu_M + \sigma^2 \, \sigma^2 \cdots \sigma^2], \quad (3.18)$$

where B is a unitary matrix of order P_1P_2 and $\mu_i > 0$, for $1 \leq i \leq M$, are the eigenvalues of ASA^H. The number M of 2-D sinusoids may be estimated by thresholding the sequence of eigenvalues. Partition the unitary matrix $B = [\mathbf{b}_1 \mathbf{b}_2 \cdots \mathbf{b}_{P_1P_2}]$ into $B = [B_S B_N]$, where $B_S = [\mathbf{b}_1 \cdots \mathbf{b}_M]$ is the matrix of eigenvectors associated with the eigenvalues, $\mu_i + \sigma^2$, for $1 \leq i \leq M$, and $B_N = [\mathbf{b}_{M+1} \cdots \mathbf{b}_{P_1P_2}]$ is the matrix of eigenvectors associated with the eigenvalues σ^2. Through the same argument as in [3], Eq. (3.18) implies that

$$A^H B_N = \mathbf{0}, \quad (3.19.a)$$

and

$$A = B_S T, \tag{3.19.b}$$

where T is a nonsingular matrix. These results will be used in all the PESS algorithms.

4. DERIVATION OF SUBSPACES FOR SEPARATE 1-D SIGNALS

4.1. Correlation Matrices and Their Estimation

$x(n_1 + k_1, n_2)$, for $0 \leq k_1 \leq P_1 - 1$, are mapped into a $P_1 \times 1$ column vector

$$\mathbf{x}_1(n_1, n_2) = [x(n_1, n_2)x(n_1 + 1, n_2) \cdots x(n_1 + P_1 - 1, n_2)]^t. \tag{4.1}$$

Substituting the measurement model of Eq. (3.1) for each element of $\mathbf{x}_1(n_1, n_2)$, we get

$$\mathbf{x}_1(n_1, n_2) = \begin{bmatrix} e^{j(\omega_{11}n_1+\omega_{21}n_2)} & \cdots & e^{j(\omega_{1M}n_1+\omega_{2M}n_2)} \\ e^{j(\omega_{11}(n_1+1)+\omega_{21}n_2)} & \cdots & e^{j(\omega_{1M}(n_1+1)+\omega_{2M}n_2)} \\ \vdots & \vdots & \vdots \\ e^{j(\omega_{11}(n_1+P_1-1)+\omega_{21}n_2)} & \cdots & e^{j(\omega_{1M}(n_1+P_1-1)+\omega_{2M}n_2)} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} + [v(n_1, n_2)\, v(n_1 + 1, n_2) \cdots v(n_1 + P_1 - 1, n_2)]^t. \tag{4.2}$$

Denoting the last term of Eq. (4.2) by $\mathbf{v}_1(n_1, n_2)$ and reorganizing the first term, we have

$$\mathbf{x}_1(n_1, n_2) = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j\omega_{11}} & \cdots & e^{j\omega_{1M}} \\ \vdots & \vdots & \vdots \\ e^{j\omega_{11}(P_1-1)} & \cdots & e^{\omega_{1M}(P_1-1)} \end{bmatrix} \cdot \begin{bmatrix} a_1 e^{j(\omega_{11}n_1+\omega_{21}n_2)} \\ a_2 e^{j(\omega_{12}n_1+\omega_{22}n_2)} \\ \vdots \\ a_M e^{j(\omega_{1M}n_1+\omega_{2M}n_2)} \end{bmatrix} + \mathbf{v}_1(n_1, n_2). \tag{4.3}$$

We know that ω_{1m}, for $1 \leq m \leq M$, may not be distinct. Suppose that ω_{1i_k}, for $1 \leq k \leq M_1$, are distinct such that the two sets, $\{\omega_{1m}\}_{m=1}^{M}$ and $\{\omega_{1i_k}\}_{k=1}^{M_1}$, are the same after ordering, and $\omega_{1i_k} = \omega_{1m}$, for $m \in \mathcal{F}_k$, where $\sum_{k=1}^{M_1}$ (number of elements in $\mathcal{F}_k$) $= M$. Note that one of the ω_{1i_k}'s may be zero. Then Eq. (4.3) can be rearranged into

$$\mathbf{x}_1(n_1, n_2) = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j\omega_{1i_1}} & \cdots & e^{j\omega_{1i_{M1}}} \\ \vdots & \vdots & \vdots \\ e^{j\omega_{1i_1}(P_1-1)} & \cdots & e^{\omega_{1i_{M1}}(P_1-1)} \end{bmatrix} \times \begin{bmatrix} \sum_{m\in\mathcal{F}_1} (a_m e^{j(\omega_{1m}n_1+\omega_{2m}n_2)}) \\ \vdots \\ \sum_{m\in\mathcal{F}_{M_1}} (a_m e^{j(\omega_{1m}n_1+\omega_{2m}n_2)}) \end{bmatrix} + \mathbf{v}_1(n_1, n_2). \tag{4.4}$$

Define $A_1 = [\mathbf{a}_1(e^{j\omega_{1i_1}}) \cdots \mathbf{a}_1(e^{j\omega_{1i_{M1}}})]$. Using $\mathbf{a}_1(\cdot)$ given by Eq. (3.9), Eq. (4.4) is written as

$$\mathbf{x}_1(n_1, n_2) = A_1 \cdot \mathbf{s}_1(n_1, n_2) + \mathbf{v}_1(n_1, n_2), \tag{4.5}$$

where $\mathbf{s}_1(n_1, n_2)$ is identified by comparing Eq. (4.4) with Eq. (4.5). The kth element of the $M_1 \times 1$ vector $\mathbf{s}_1(n_1, n_2)$ can be rewritten as $e^{j\omega_{1i_k}n_1} \cdot \sum_{m\in\mathcal{F}_k} a_m e^{j\omega_{2m}n_2}$. Also the elements of the set $\{\omega_{1i_k}\}$, for $1 \leq k \leq M_1$, are distinct. Thus, using similar arguments as in Sec. 3.1, we can conclude that asymptotically, $\mathbf{s}_1(n_1, n_2)$ is a random vector with a nonsingular correlation matrix

$$S_1 = E[\mathbf{s}_1(n_1, n_2)\mathbf{s}_1^H(n_1, n_2)].$$

The correlation matrix R_1 of $\mathbf{x}_1(n_1, n_2)$ has a structure similar to that of R in Eq. (3.12).

$$R_1 = E[\mathbf{x}_1(n_1, n_2)\mathbf{x}_1^H(n_1, n_2)] = A_1 S_1 A_1^H + \sigma^2 I. \tag{4.6}$$

In Eq. (4.6), A_1 and S_1 are the counterparts of A and S in Eqs. (3.6) and (3.11). Actually, $\mathbf{x}_1(n_1, n_2)$ is a window through which we look at the data matrix $\mathbf{X}$ of Eq. (2.2). The window of size $P_1 \times 1$ moves horizontally with indexing n_2 and vertically with indexing n_1 within the data matrix X in Eq. (2.2). $\mathbf{x}_1(n_1, n_2)$, for $0 \leq n_1 \leq N_1 - P_1$ and $0 \leq n_2 \leq N_2$, are interpreted to be the realizations of the random process $\mathbf{x}_1(n_1, n_2)$. Define the reorganized data matrix X_{r1} by

$$X_{r1} = \begin{matrix} [\mathbf{x}_1(0,0) & \mathbf{x}_1(1,0) \\ \mathbf{x}_1(0,1) & \mathbf{x}_1(1,1) \\ \cdots & \cdots \\ \cdots & \cdots \\ \mathbf{x}_1(0,N_2-1) & \mathbf{x}_1(1,N_2-1) \end{matrix} \begin{matrix} \cdots & \mathbf{x}_1(N_1-P_1,0) \\ \cdots & \mathbf{x}_1(N_1-P_1,1) \\ \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \mathbf{x}_1(N_1-P_1,N_2-1)] \end{matrix}, \tag{4.7}$$

which has the order $P_1 \times ((N_1 - P_1 + 1)N_2)$. Assuming correlation-ergodicity of $\mathbf{x}_1(n_1, n_2)$, the estimate $\hat{R}_1$ of the correlation matrix R_1 is computed by

$$\hat{R}_1 = \frac{1}{2(N_1 - P_1 + 1)N_2}(X_{r1}X_{r1}^H + JX_{r1}^*X_{r1}^t J). \tag{4.8}$$

Next, the data elements $x(n_1, n_2 + k_2)$, for $0 \leq k_2 \leq P_2 - 1$, are stacked into a $P_2 \times 1$ column vector

$$\mathbf{x}_2(n_1, n_2) = [x(n_1, n_2)x(n_1, n_2 + 1) \cdots x(n_1, n_2 + P_2 - 1)]^t. \quad (4.9)$$

Through the same argument as in the case of $\mathbf{x}_1(n_1, n_2)$, we get

$$\mathbf{x}_2(n_1, n_2) = A_2 \cdot \mathbf{s}_2(n_1, n_2) + \mathbf{v}_2(n_1, n_2), \quad (4.10)$$

where $A_2 = [\mathbf{a}_2(e^{j\omega_{21}}) \cdots \mathbf{a}_2(e^{j\omega_{2M_2}})]$, M_2 is the number of distinct ω_{2i}'s, and S_2 defined by $E[\mathbf{s}_2(n_1, n_2)\mathbf{s}_2^H(n_1, n_2)]$ is nonsingular. The correlation matrix R_2 of $\mathbf{x}_2(n_1, n_2)$ has the structure expressed by

$$R_2 = E[\mathbf{x}_2(n_1, n_2)\mathbf{x}_2^H(n_1, n_2)] = A_2 S_2 A_2^H + \sigma^2 I, \quad (4.11)$$

where A_2 and S_2 are defined analogously to A_1, S_1, or A, S. $\mathbf{x}_2^t(n_1, n_2)$ is a window through which we look at the data matrix X of Eq. (2.2). The window of size $1 \times P_2$ moves horizontally and vertically with indexing n_2 and n_1, respectively. Define the reorganized data matrix X_{r2} by

$$X_{r2} = \begin{matrix} [\mathbf{x}_2(0,0) & \mathbf{x}_2(0,1)) \\ \mathbf{x}_2(1,0) & \mathbf{x}_2(1,1) \\ \cdots & \cdots \\ \cdots & \cdots \\ \mathbf{x}_2(N_1-1,0) & \mathbf{x}_2(N_1-1,1) \end{matrix} \begin{matrix} \cdots & \mathbf{x}_2(0,N_2-P_2) \\ \cdots & \mathbf{x}_2(1,N_2-P_2) \\ \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \mathbf{x}_2(N_1-1,N_2-P_2)] \end{matrix}, \quad (4.12)$$

which is of order $P_2 \times (N_1(N_2 - P_2 + 1))$. Assuming correlation-ergodicity of $\mathbf{x}_2(n_1, n_2)$, the estimate $\hat{R}_2$ of the correlation matrix R_2 is computed by

$$\hat{R}_2 = \frac{1}{2N_1(N_2 - P_2 + 1)}(X_{r2}X_{r2}^H + JX_{r2}^*X_{r2}^t J). \quad (4.13)$$

4.2. Derivation of Subspaces

Since the correlation matrices R_1 and R_2 are Hermitian, and $A_1S_1A_1^H$ and $A_2S_2A_2^H$ are of rank M_1 and M_2, respectively, the eigendecompositions of R_1 and R_2 have the structures described by

$$B_1^H R_1 B_1 = B_1^H A_1 S_1 A_1^H B_1 + \sigma^2 I = \text{Diag}[\mu_{11} + \sigma_2 \cdots \mu_{1M_1} + \sigma^2\ \sigma^2 \cdots \sigma^2], \quad (4.14.a)$$

$$B_2^H R_2 B_2 = B_2^H A_2 S_2 A_2^H B_2 + \sigma^2 I = \text{Diag}[\mu_{21} + \sigma^2 \cdots \mu_{2M_2} + \sigma^2\ \sigma^2 \cdots \sigma^2], \quad (4.14.b)$$

where B_1 and B_2 are unitary matrices, and μ_{1i}, for $1 \leq i \leq M_1$, and μ_{2i}, for $1 \leq i \leq M_2$, are positive. The numbers, M_1 and M_2, of distinct 1-D parameters may be estimated by thresholding the sequences of eigenvalues. We know that $1 \leq M_1 \leq M$ and $1 \leq M_2 \leq M$. In PESS, M_1 and M_2 are not required. Partition the unitary matrices $B_1 = [\mathbf{b}_{11}\mathbf{b}_{12} \cdots \mathbf{b}_{1P_1}]$ and $B_2 = [\mathbf{b}_{21}\mathbf{b}_{22} \cdots \mathbf{b}_{2P_2}]$ into $B_1 = [B_{1S}B_{1N}]$ and $B_2 = [B_{2S}B_{2N}]$, respectively, where $B_{1S} = [\mathbf{b}_{11} \cdots \mathbf{b}_{1M_1}]$ and $B_{2S} = [\mathbf{b}_{21} \cdots \mathbf{b}_{2M_2}]$ are the matrices of eigenvectors associated with the eigenvalues, $\mu_{1i} + \sigma^2$, for $1 \leq i \leq M_1$, and $\mu_{2i} + \sigma^2$, for $1 \leq i \leq M_2$, respectively. $B_{1N} = [\mathbf{b}_{1(M_1+1)} \cdots \mathbf{b}_{1P_1}]$ and $B_{2N} = [\mathbf{b}_{2(M_2+1)} \cdots \mathbf{b}_{2P_2}]$ are the matrices of eigenvectors associated with the eigenvalues σ^2 of R_1 and R_2, respectively. Using the argument in [3], Eqs. (4.14.a) and (4.14.b) imply that

$$A_1^H B_{1N} = 0, \quad (4.15.a)$$

$$A_1 = B_{1S}T_1, \quad (4.15.b)$$

$$A_2^H B_{2N} = 0, \quad (4.16.a)$$

and

$$A_2 = B_{2S}T_2, \quad (4.16.b)$$

where T_1 and T_2 are nonsingular matrices. These results will be used only in the 1-D based 2-D PESS.

5. DERIVATION OF THE 1-D BASED 2-D PESS METHOD

5.1. Interpretation of the Noise Subspaces

Using Eq. (3.19.a), we can easily set up the straightforward 2-D extension of the MUSIC algorithm as follows. Find (ω_1, ω_2)'s where the peaks of the function below occur.

$$P(\omega_1, \omega_2) = \frac{1}{\mathbf{a}^H(e^{j\omega_1}, e^{j\omega_2})B_N B_N^H \mathbf{a}(e^{j\omega_1}, e^{j\omega_2})}. \quad (5.1)$$

This algorithm suffers from prohibitive computational load, and the 2-D generalization of the ROOT-MUSIC algorithm is not routine because the zeros of bivariate polynomials trace continuous algebraic curves [1], and the straightforward extension of the Vandermonde system to the 2-D situation cannot be derived. In this section, an indirect 2-D generalization of the Vandermonde system is defined and, based on it, a new method for estimating the wavenumber parameters, $(\omega_{1i}, \omega_{2i})$, for $1 \leq i \leq M$, is developed.

For each column $\mathbf{b}_{1i}$ of B_{1N} given by Eq. (4.15.a), construct polynomial equations

$$\mathbf{b}_{1i}^H \mathbf{a}_1(z_1) = 0, \tag{5.2.a}$$

and, for $M_1 + 1 \leq i \leq P_1$,

$$J\mathbf{b}_{1i}\mathbf{a}_1(z_1) = 0, \tag{5.2.b}$$

where $\mathbf{a}_1(z_1) = [1\ z_1 \cdots z_1^{P_1-1}]^t$. Eq. (5.2.a) and (5.2.b), for $M_1 + 1 \leq i \leq P_1$, include the separate 1-D signal zeros, $e^{j\omega_{1i_k}}$, for $1 \leq k \leq M_1$, one of which may be $e^{j\cdot 0}$, as implied by Eq. (4.15.a). Using Eqs. (5.2.a) and (5.2.b), construct the set of $(P_1 - M_1)$ polynomial equations each of degree P_1, as shown below.

$$\sum_{k=0}^{P_1} t_{1k}^{(i-M_1)} \cdot z_1^k \triangleq z_1 \cdot (\mathbf{b}_{1i}^H + J\mathbf{b}_{1i})\mathbf{a}_1(z_1) = 0, \quad \text{for } M_1 + 1 \leq i \leq P_1. \tag{5.3}$$

Let the resulting set of $(P_1 - M_1)$ polynomial equations be described by

$$\sum_{k=0}^{P_1} t_{1k}^{(i)} z_1^k = 0, \quad 1 \leq i \leq P_1 - M_1 + 1. \tag{5.4}$$

The $(P_1 - M_1)$ sets of coefficients, $\{t_{1k}^{(i)}\}_{k=1}^{P_1}$, for $1 \leq i \leq P_1 - M_1 + 1$, are interpreted as the statistical data set for a generic set of coefficients, denoted by $\{t_{1k}\}_{k=1}^{P_1}$. Without loss of generality, we can assume that $t_{1P_1} \neq 0$. If $t_{1k} = 0$, for $m \leq k \leq P_1$ and $t_{1(m-1)} \neq 0$, then we multiply both sides of Eq. (5.4) by $z_1^{P_1-m+1}$. Manipulate Eqs. (4.16.a) and (4.16.b) in a similar way to get a corresponding set of equations, given next, in the other dimension.

$$\mathbf{b}_{2i}^H \mathbf{a}_2(z_2) = 0, \tag{5.5.a}$$

$$J\mathbf{b}_{2i}\mathbf{a}_2(z_2) = 0, \tag{5.5.b}$$

and, for $M_2 + 1 \leq i \leq P_2$,

$$\sum_{k=0}^{P_2} t_{2k}^{(i-M_2)} \cdot z_2^k \triangleq z_2 \cdot (\mathbf{b}_{2i}^H + J\mathbf{b}_{2i})\mathbf{a}_2(z_2) = 0. \tag{5.6}$$

Finally, we have a generic polynomial equation representing Eq. (5.6), as given below.

$$\sum_{k=0}^{P_2} t_{2k} z_2^k = 0, \quad \text{with } t_{2P_2} \neq 0. \tag{5.7}$$

We know that Eq. (5.4) and Eq. (5.7) include the separate 1-D signal zeros, $e^{j\omega_{1i_k}}$, for $1 \leq k \leq M_1$, and $e^{j\omega_{2i_r}}$, for $1 \leq r \leq M_2$, respectively. When M_1 and M_2 are not available, M replaces those. In simulation, averages of the coefficient sets were substituted in the generic polynomial equations.

5.2. A 2-D Vandermonde System

A 2-D Vandermonde system is defined, based on a set of 1-D Vandermonde system given below. Eq. (5.4) and (5.7) can be reformulated as 1-D Vandermonde systems expressed by

$$D_1 \cdot \mathbf{a}_1(z_1) = \mathbf{a}_1(z_1) \cdot z_1 \tag{5.8.a}$$

$$D_2 \cdot \mathbf{a}_2(z_2) = \mathbf{a}_2(z_2) \cdot z_2 \tag{5.8.b}$$

where the companion matrices, for $m = 1$ and 2, are

$$D_m = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\dfrac{t_{m,0}}{t_{m,P_m}} & -\dfrac{t_{m,1}}{t_{m,P_m}} & -\dfrac{t_{m,2}}{t_{m,P_m}} & \cdots & -\dfrac{t_{m,P_m-1}}{t_{m,P_m}} \end{bmatrix}. \tag{5.8.c}$$

Taking the Kronecker products of the left-hand sides and the right-hand sides of Eq. (5.8.a) and (5.8.b), we get

$$(D_2 \otimes D_1) \cdot [\mathbf{a}_2(z_2) \otimes \mathbf{a}_1(z_1)] = [\mathbf{a}_2(z_2) \otimes \mathbf{a}_1(z_1)] \cdot (z_1 z_2), \tag{5.9}$$

which will be called a *2-D Kronecker product Vandermonde system.* From the property of the Kronecker product, we know that $e^{j\omega_{1i}} \cdot e^{j\omega_{2i}}$ is the eigenvalue associated with $\mathbf{a}_2(e^{j\omega_{2i}}) \otimes \mathbf{a}_1(e^{j\omega_{1i}})$, which is an eigenvector of $(D_2 \otimes D_1)$, for $1 \leq i \leq M$. For arbitrary complex numbers g_1 and g_2, Eqs. (5.8.a) and (5.8.b) can be rewritten as

$$(g_1 D_1) \cdot \mathbf{a}_1(z_1) = \mathbf{a}_1(z_1) \cdot (g_1 z_1) \tag{5.10.a}$$

$$(g_2 D_2) \cdot \mathbf{a}_2(z_2) = \mathbf{a}_2(z_2) \cdot (g_2 z_2). \tag{5.10.b}$$

From a property of the Kronecker sum denoted by $\oplus$, ($B \oplus C = B \otimes I_C + I_B \otimes C$, where I_C and I_B are the identity matrices of the same order as C and B, respectively) it follows that

$$[(g_2 D_2) \oplus (g_1 D_1)] \cdot [\mathbf{a}_2(z_2) \otimes \mathbf{a}_1(z_1)] = [\mathbf{a}_2(z_2) \otimes \mathbf{a}_1(z_1)] \cdot (g_1 z_1 + g_2 z_2). \tag{5.11}$$

Equation (5.11) implies that $g_1 e^{j\omega_{1i}} + g_2 e^{j\omega_{2i}}$ is the eigenvalue associated with $\mathbf{a}_2(e^{j\omega_{2i}}) \otimes \mathbf{a}_1(e^{j\omega_{1i}})$, which is an eigenvector of $[(g_2 D_2) \oplus (g_1 D_1)]$, for $1 \leq i \leq M$, where either g_1 or g_2 must be non-zero. Equation (5.11) is called a *2-D Kronecker sum Vandermonde system.*

Multiplying both sides of Eq. (5.9) by a complex number g_3 and adding the resulting equation to Eq. (5.11), we have

$$[g_3(D_2 \otimes D_1) + \{(g_2D_2) \oplus (g_1D_1)\}] \cdot [\mathbf{a}_2(z_2) \otimes \mathbf{a}_1(z_1)] = [\mathbf{a}_2(z_2) \otimes \mathbf{a}_1(z_1)] \cdot (g_3z_1z_2 + g_2z_2 + g_1z_1), \quad (5.12)$$

which can be expressed by

$$D(\mathbf{g}) \cdot \mathbf{a}(z_1, z_2) = \lambda(\mathbf{g}) \cdot \mathbf{a}(z_1, z_2), \quad (5.13.a)$$

where

$$D(\mathbf{g}) = g_3(D_2 \otimes D_1) + \{(g_2D_2) \oplus (g_1D_1)\} = g_3(D_2 \otimes D_1) + g_2(D_2 \otimes I_{P_1}) + g_1(I_{P_2} \otimes D_1), \quad (5.13.b)$$

$$\lambda(\mathbf{g}) = g_3z_1z_2 + g_2z_2 + g_1z_1, \quad (5.13.c)$$

and $\mathbf{g} = [g_1, g_2, g_3]^t$ is a non-zero vector (I_i denotes the identity matrix of order i). Eq. (5.13a) is called a *2-D Vandermonde system.* We know that $g_3e^{j\omega_{1i}} \cdot e^{j\omega_{2i}} + g_2e^{j\omega_{2i}} + g_1e^{j\omega_{1i}}$ is the eigenvalue associated with $\mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}})$, for $1 \leq i \leq M$, which, for any non-zero complex vector $\mathbf{g}$, is an eigenvector of $D(\mathbf{g})$. Similarly, the M eigenvalues are denoted by $\lambda_i(\mathbf{g})$, for $1 \leq i \leq M$. Then from the definition of the matrix A and Eq. (5.13.b), we have, for any nonzero $\mathbf{g}$,

$$D(\mathbf{g}) \cdot A = A \cdot \Lambda(\mathbf{g}), \quad (5.14)$$

where $\Lambda(\mathbf{g}) = \text{Diag}[\lambda_1(\mathbf{g}) \cdots \lambda_M(\mathbf{g})]$ and $\lambda_i(\mathbf{g}) = g_3e^{j\omega_{1i}} \cdot e^{j\omega_{2i}} + g_2e^{j\omega_{2i}} + g_1e^{j\omega_{1i}}$.

5.3. Obtaining the Parameter Vectors

Using Eqs. (3.19.b) and (5.14), the 2-D parameter vectors, $(\omega_{1i}, \omega_{2i})$, for $1 \leq i \leq M$, are obtained from the results of the following theorem, which shows the signal-selectivity of signal subspace. The theorem is formulated and proved in a general framework.

THEOREM 5.1. *Consider matrices, D of order $p \times p$, A and E each of order $p \times q$ such that $p > q$, let E be of full rank,* Span*(A)* = Span*(E), and $DA = A\Lambda$, where Λ is a $(q \times q)$ diagonal matrix. Then, the following eigendecomposition exists*

$$[(E^HE)^{-1}E^HDE] = T\Delta T^{-1}, \quad (5.15)$$

where Δ is the same as Λ after column or row permutation. Furthermore, for each eigenvalue λ of multiplicity m in Δ,

$$\text{Span}(A_\lambda) = \text{Span}(ET_\lambda), \quad (5.16)$$

where A_λ and T_λ are the matrices of m eigenvectors associated with λ in A and T, respectively.

Proof. After substituting $A = ET$, where T should be nonsingular matrix, for A in $DA = A\Lambda$, the least-squares solution for $T\Lambda T^{-1}$ is obtained as $(E^HE)^{-1}E^HDE = T\Lambda T^{-1}$. This implies that $(E^HE)^{-1}e^HDE$ is diagonalizable. Pre-multiplying and post-multiplying both sides of Eq. (5.15) by E and T, respectively, we get, $E(E^HE)^{-1}E^HDET = ET\Delta$. Since Span$(ET)$ = Span(A), therefore, $ET = AW$ for some nonsingular matrix W. Therefore, $E(E^HE)^{-1}E^HDAW = AW\Delta$, which reduces to $E(E^HE)^{-1}E^HA\Lambda W = AW\Delta$, since $DA = A\Lambda$. From the fact that $A\Lambda W \in$ Span(E) and $E(E^HE)^{-1}E^H$ is the projection operator to Span(E), it follows that $A\Lambda W = AW\Delta$. Since A is also of full rank, therefore, $\Lambda W = W\Delta$, or $\Lambda = W\Delta W^{-1}$. Since Δ and Λ are diagonal matrices, and the last equation represents an eigendecomposition, therefore, $\Delta = \Lambda$ within permutation.

Suppose, for convenience of argument, that $\Delta = \Lambda$, and λ_i, for $1 \leq i \leq r(\leq q)$, are the distinct elements in Λ, and λ_i is of multiplicity m_i. Then, it follows that $\Lambda = W\Lambda W^{-1}$, where $\Lambda = \text{Diag}[\text{Diag}[\lambda_1 \cdots \lambda_1] \cdots \text{Diag}[\lambda_r \cdots \lambda_r]]$ and $W = \text{Diag}[W_1W_2 \cdots W_r]$, where the W_i's are nonsingular matrices. Thus, from $ET = AW$, it follows that $A = ETW^{-1}$, or $A_{\lambda_i} = ET_{\lambda_i}W_i^{-1}$, for $1 \leq i \leq r$. ■

The matrices, A, B_S, and D given by Eq. (3.19.b) and (5.14) satisfy the assumption of Theorem 5.1. If we have an eigendecomposition,

$$(B_S^HB_S)^{-1}B_S^HD(\mathbf{g})B_S = T(\mathbf{g})\Delta(\mathbf{g})T(\mathbf{g})^{-1}, \quad (5.17)$$

where $\Delta(\mathbf{g})$ has the eigenvalues $\lambda_i(\mathbf{g})$ of multiple m_i, for $1 \leq i \leq r$ with $r \leq M$, and $T_i(\mathbf{g})$ is the matrix of all the eigenvectors corresponding to $\lambda_i(\mathbf{g})$, then we have that

$$\text{Span}(A_{\lambda_i(\mathbf{g})}) = \text{Span}(B_ST_i(\mathbf{g})), \quad \text{for } 1 \leq i \leq r. \quad (5.18)$$

If $m_i = 1$, then Eq. (5.18) implies that

$$A_{\lambda_i(\mathbf{g})} = B_ST_i(\mathbf{g}) \cdot \alpha_i, \quad (5.19)$$

where $T_i(\mathbf{g})$ is actually a column vector, $A_{\lambda_i(\mathbf{g})}$ is actually $\mathbf{a}(e^{j\omega_{1k}}, e^{j\omega_{2k}})$, for some k, and α_i is a scalar. When we scale $B_ST_i(\mathbf{g})$ so that its first element is 1, let the second element be $e^{j\omega_{1k}}$ and the $(P_1 + 1)$th element be $e^{j\omega_{2k}}$. Then $(\omega_{1k}, \omega_{2k})$ is a desired parameter vector. If $m_i = 1$ for all i, then Eq. (5.18) implies that

$$A = B_ST(\mathbf{g}) \cdot \Gamma, \quad (5.20)$$

where Γ is a nonsingular diagonal matrix. When the scaling is done so that the first row of $B_ST(\mathbf{g})$ is

$[1 \cdots 1]$, the M parameter vectors, $(\omega_{1i}, \omega_{2i})$, for $1 \leq i \leq M$, can be easily identified from Eq. (5.20).

If $M_i > 1$, we apply the Theorem 5.1 again. The matrices, $A_{\lambda_i(\mathbf{g})}$, $B_S T_i(\mathbf{g})$, and $D(\mathbf{g}')$ given by Eq. (5.18) and Eq. (5.14) satisfy the assumption of Theorem 5.1. Note that $\mathbf{g}'$ is different from $\mathbf{g}$, so that the eigenvalues of

$$[T_i(\mathbf{g})^H B_S^H B_S T_i(\mathbf{g})]^{-1} T_i(\mathbf{g})^H B_S^H D(\mathbf{g}') B_S T_i(\mathbf{g}) \quad (5.21)$$

are distinct. Denote by $T_i(\mathbf{g}, \mathbf{g}')$ the $m_i \times m_i$ eigenvector matrix of Eq. (5.21), then, we have

$$A_{\lambda_i(\mathbf{g})} = B_S T_i(\mathbf{g}, \mathbf{g}') \cdot \Gamma_i, \quad (5.22)$$

where Γ_i is a non-singular matrix. Thus, $A_{\lambda_i(\mathbf{g})}$ is determined uniquely by selecting Γ_i so that the first row of Eq. (5.22) is $[1 \cdots 1]$.

When the estimate $\hat{A}$ of A has been obtained, each column of $\hat{A}$ represents a paired parameter vector. In order to compute the parameters, the structure of A is exploited. We know that

$$A = \begin{bmatrix} 1 \\ 1 \cdot \Phi_1 \\ \vdots \\ 1 \cdot \Phi_1^{P_1-1} \\ 1 \cdot \Phi_2 \\ 1 \cdot \Phi_2 \cdot \Phi_1 \\ \vdots \\ 1 \cdot \Phi_2 \cdot \Phi_1^{P_1-1} \\ \vdots \\ 1 \cdot \Phi_2^{P_2-1} \\ 1 \cdot \Phi_2^{P_2-1} \cdot \Phi_1 \\ \vdots \\ 1 \cdot \Phi_2^{P_2-1} \cdot \Phi_1^{P_1-1} \end{bmatrix}, \quad (5.23)$$

where $\Phi_1 = \text{Diag}[e^{j\omega_{11}} \cdots e^{j\omega_{1M}}]$ and $\Phi_2 = \text{Diag}[e^{j\omega_{21}} \cdots e^{j\omega_{2M}}]$. Let $\hat{A}$ be defined by the column vectors named below.

$$\hat{A} \triangleq [\mathbf{s}_{1,1}^t \mathbf{s}_{2,1}^t \cdots \mathbf{s}_{P_1-1,1}^t \mathbf{s}_{1,2}^t \mathbf{s}_{2,2}^t \cdots \mathbf{s}_{P_1-1,2}^t \cdots \mathbf{s}_{1,P_2-1}^t \mathbf{s}_{2,P_2-1}^t \cdots \mathbf{s}_{P_1-1,P_2-1}^t]^t. \quad (5.24)$$

Then it follows that

$$S_{1a}\Phi_1 = S_{1b} \quad \text{and} \quad S_{2a}\Phi_2 = S_{2b}, \quad (5.25)$$

where

$$S_{1a} = [\mathbf{s}_{1,1}^t \mathbf{s}_{2,1}^t \cdots \mathbf{s}_{P_1-2,1}^t \cdots \mathbf{s}_{1,P_2-1}^t \mathbf{s}_{2,P_2-1}^t \cdots \mathbf{s}_{P_1-2,P_2-1}^t]^t,$$

$$S_{1b} = [\mathbf{s}_{2,1}^t \mathbf{s}_{3,1}^t \cdots \mathbf{s}_{P_1-1,1}^t \cdots \mathbf{s}_{2,P_2-1}^t \mathbf{s}_{3,P_2-1}^t \cdots \mathbf{s}_{P_1-1,P_2-1}^t]^t$$

$$S_{2a} = [\mathbf{s}_{1,1}^t \mathbf{s}_{2,1}^t \cdots \mathbf{s}_{P_1-1,1}^t \cdots \mathbf{s}_{1,P_2-1}^t \mathbf{s}_{2,P_2-2}^t \cdots \mathbf{s}_{P_1-1,P_2-2}^t]^t,$$

and

$$S_{2b} = [\mathbf{s}_{1,2}^t \mathbf{s}_{2,2}^t \cdots \mathbf{s}_{P_1-1,2}^t \cdots \mathbf{s}_{1,P_2-1}^t \mathbf{s}_{2,P_2-1}^t \cdots \mathbf{s}_{P_1-1,P_2-1}^t]^t.$$

Let the least squares or total least squares solution [20, p. 222] of Eq. (5.25) be $\hat{\Phi}_1$ and $\hat{\Phi}_2$, then the ith diagonal elements of $\hat{\Phi}_1$ and $\hat{\Phi}_2$ represent the two elements of the ith parameter vector, for $1 \leq i \leq M$.

5.4. Selection of g

Define column vectors, $\mathbf{e}_i^t = [e^{j\omega_{1i}}, e^{j\omega_{2i}}, e^{j(\omega_{1i}+\omega_{2i})}]$, for $1 \leq i \leq M$. Then $\lambda_i(\mathbf{g}) = \mathbf{g}^t\mathbf{e}_i$, for $1 \leq i \leq M$, are the eigenvalues of $(B_S^H B_S)^{-1} B_S^H D(\mathbf{g}) B_S$. Since $(e^{j\omega_{1i}}, e^{j\omega_{2i}})$, for $1 \leq i \leq M$, are distinct by assumption, $\mathbf{e}_i^t - \mathbf{e}_k^t \neq 0$ whenever $i \neq k$. In order that $\lambda_i(\mathbf{g})$, for $1 \leq i \leq M$, are distinct, it must follow that $\mathbf{g}^t\mathbf{e}_i \neq \mathbf{g}^t\mathbf{e}_k$ or $\mathbf{g}^t(\mathbf{e}_i - \mathbf{e}_k) \neq 0$, for any i, k with $i \neq k$, which implies that $\mathbf{g}^*$ (the superscript * denotes complex conjugate) must not be orthogonal to $\mathbf{e}_i - \mathbf{e}_k$ for $1 \leq i, k \leq M$ and $i \neq k$, since $\mathbf{e}_i - \mathbf{e}_k \neq \mathbf{0}$ whenever $i \neq k$. That is, Δ in Eq. (5.15) has only distinct eigenvalues if $\mathbf{g}$ does not belong to the $M(M-1)/4$ hyperplanes expressed by $\mathbf{g}^t(\mathbf{e}_i - \mathbf{e}_k) = 0$ for $1 \leq i, k \leq M$ and $i \neq k$. Thus, if $\mathbf{g}$ is selected randomly, the probability that Δ in Eq. (5.15) has non-distinct eigenvalues is, theoretically, zero. (In computation, we must consider some numerical margin.) It is good to select $\mathbf{g}$ randomly. Then $\mathbf{g}'$ is selected to be orthogonal to $\mathbf{g}$, i.e., $(\mathbf{g}')^H\mathbf{g} = 0$. If $\mathbf{e}_i - \mathbf{e}_k$ is orthogonal to $\mathbf{g}^*$, then it must be not orthogonal to $(\mathbf{g}')^*$. Thus, $\lambda_i(\mathbf{g}')$ and $\lambda_k(\mathbf{g}')$ become different although $\lambda_i(\mathbf{g})$ and $\lambda_k(\mathbf{g})$ were the same.

Considering the fact that, if $\mathbf{g}$ is selected randomly, the probability that Δ in Eq. (5.15) has non-distinct eigenvalues is zero theoretically, we can establish another algorithm. First, for $L(\geq 2)$ randomly generated $[g_1\, g_2\, g_3]^t$s, solve the Eq. (5.17) and obtain A's by $A = B_S T(\mathbf{g})$. Second, choose A which either minimizes [7,8]

$$S_N = \sum_{i=1}^{M} \mathbf{a}^H(e^{j\omega_{1i}}, e^{j\omega_{2i}}) B_N B_N^H \mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) \quad (5.26)$$

or maximizes

$$S_S = \sum_{i=1}^{M} \mathbf{a}^H(e^{j\omega_{1i}}, e^{j\omega_{2i}}) B_S B_S^H \mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) \quad (5.27)$$

5.5. The Algorithms

According to the discussion for $\mathbf{g}$ in Section 5.4, we can set up three algorithms for estimating 2-D parameter vectors. Algorithm 1 tests the magnitude of the difference (referred to, henceforth, as distance) between two complex-valued eigenvalues to find if there are non-distinct eigenvalues in the matrix Δ of Eq.

(5.15). Algorithm 2 randomizes the set of eigenvalues and finds distinct eigenvalues by calculating the distances between the eigenvalues. Algorithm 3 randomizes the set of eigenvalues and finds the set of 2-D parameter vectors by examining their fitness obtained by evaluating the objective function of Eq. (5.26) or (5.27).

In the following, it is assumed that a sample data matrix X has been given as in Eq. (2.2), the number M of 2-D wavenumbers is known, and $P_1(>M)$ and $P_2(>M)$ have been chosen.

5.5.1. *The 1-D Based 2-D PESS Method with Distance Test of Eigenvalues*

Step 1. Construct X_r from the sample data matrix X by Eqs. (3.13) and (3.15). Compute the correlation matrix $\hat{R}$ by Eq. (3.16). Compute the eigendecomposition of $\hat{R}$ to get $\hat{R} = \hat{B}\hat{\Delta}\hat{B}^H$. Partition $\hat{B}$ as $\hat{B} = [\hat{B}_N \hat{B}_S]$, where $\hat{B}_N$ is the $(P_1P_2) \times (P_1P_2 - M)$ matrix of the eigenvectors associated with the smallest $(P_1P_2 - M)$ eigenvalues. The sequence of eigenvalues in Δ may be tested to estimate the number of 2-D wavenumbers.

Step 2.a. Construct X_{r1} and X_{r2} from the sample data matrix X by Eqs. (4.1), (4.7), (4.9), and (4.12). Compute the correlation matrices $\hat{R}_1$ and $\hat{R}_2$ by Eqs. (4.8) and (4.13). Compute the eigendecompositions of $\hat{R}_i$ to get $\hat{R}_i = \hat{B}_i\hat{\Delta}_i\hat{B}_i^H$. Partition $\hat{B}_i$ as $\hat{B}_i = [\hat{B}_{iN}\hat{B}_{iS}]$, where $\hat{B}_{iN}$ is the $P_i \times (P_i - M)$ matrix of the eigenvectors associated with the smallest $(P_i - M)$ eigenvalues. Two univariate polynomial equations in z_i, for $i = 1, 2$, are selected as

$$\{(\hat{B}_{iN} + J\hat{B}_{iN}^H) \cdot \mathbf{1}\}^H \mathbf{a}_i(z_i) \cdot z_i = 0,$$

where $\mathbf{1}$ is the column vector of appropriate order, whose elements are all 1.

Step 2.b. Construct two companion matrices D_2 and D_1 by Eqs. (5.8.a), (5.8.b), (5.8.c). Compute

$$G_1 = (\hat{B}_S^H\hat{B}_S)^{-1}\hat{B}_S^H(I \otimes D_1),$$

where I is the identity matrix of order P_2,

$$G_2 = (\hat{B}_S^H\hat{B}_S)^{-1}\hat{B}_S^H(D_1 \otimes I),$$

where I is the identity matrix of order P_1 and

$$G_3 = (\hat{B}_S^H\hat{B}_S)^{-1}\hat{B}_S^H(D_2 \otimes D_1).$$

Step 3. Select $\mathbf{g}^{(1)}(\neq \mathbf{0})$ and $\mathbf{g}^{(2)}(\neq \mathbf{0})$ such that they are orthogonal. Also select a threshold value δ.

Step 4. Compute $F = g_1^{(1)}G_1 + g_2^{(1)}G_2 + g_3^{(1)}G_3$ with $\mathbf{g}^{(1)}$. Compute the eigendecomposition of F to get $F = T \cdot \mathrm{Diag}[\lambda_1 \cdots \lambda_M] \cdot T^{-1}$, where the eigenvector matrix is $T \triangleq [\mathbf{t}_1 \cdots \mathbf{t}_M]$.

Step 5. Compute $\mathrm{dist}_{i,j} = |\lambda_i - \lambda_j|$, for $1 \leq i, j \leq M$, and $i > j$. Construct the clusters $\mathcal{C}_i$ of eigenvalues such that $|\alpha - \beta| > \delta$ for $\alpha \in \mathcal{C}_k$ and $\beta \in \mathcal{C}_j$ whenever $k \neq j$ and $|\alpha - \beta| \leq \delta$ for $\alpha, \beta \in \mathcal{C}_k$. Suppose that the number of clusters is n.

Step 6. If the ith cluster has one member, keep $\mathbf{t}_i$ unchanged, where $\mathbf{t}_i$ is the eigenvector associated with the eigenvalue in the ith cluster.

Step 7. If the ith cluster has m members, compute $F_i = (T_i^HT_i)^{-1}T_i^H(g_1^{(2)}G_1 + g_2^{(2)}G_2 + g_3^{(2)}G_3)T_i$ with $\mathbf{g}^{(2)}$, where T_i is the matrix of the eigenvectors associated with the eigenvalues in the ith cluster. Compute the eigenvector matrix U_i of F_i and replace T_i by T_iU_i.

Step 8. Compute the estimate $\hat{A}$ of A as $\hat{A} = \hat{B}_ST$. Construct S_{1a}, S_{1b}, S_{2a}, S_{2b} by Eq. (5.25). Solve the linear systems via the method of least squares to get

$$\Phi_1 = (S_{1a}^HS_{1a})^{-1}S_{1a}^HS_{1b}$$

and

$$\Phi_2 = (S_{2a}^HS_{2a})^{-1}S_{2a}^HS_{2b}.$$

$\{(\Phi_1)_{ii}, (\Phi_2)_{ii})\}_{i=1}^M$ is the set of 2-D signal zeros which give the estimate of 2-D wavenumbers.

5.5.2. *The 1-D Based 2-D PESS Method with Distance Test of Randomized Eigenvalues*

Step 1, 2. Same as that of the algorithm in Subsection 5.5.1.

Step 3. Initialize outer loop; $m = 1$, $G_1^{(1)} = G_1$, $G_2^{(1)} = G_2$, $G_3^{(1)} = G_3$ and $D = I$, where I is the identity matrix of order M. Choose r_{0i} and r_{1i}, for $1 \leq i \leq 3$, in $g_i = r_ie^{j\phi_i}$ with $r_{0i} \leq r_i \leq r_{1i}$.

Step 4. Initialize inner loop; $n = 0$, $d_{\max} = -1$.

Step 5. Generate a complex random vector (g_1, g_2, g_3). Compute $F = g_1G_1^{(m)} + g_2G_2^{(m)} + g_3G_3^{(m)}$. Compute the eigendecomposition $F = U \cdot \mathrm{Diag}[\lambda_1 \cdots \lambda_M] \cdot U^{-1}$.

Step 6. Compute

$$d = \max_{1 \leq i \leq M} \min_{1 \leq k \leq M, k \neq i} |\lambda_i - \lambda_k|$$

and denote the maximizing argument of i by q. If $d > d_{\max}$ then set $d_{\max} = d$, $q_{\max} = q$ and $T = U$.

Step 7. Compute $n = n + 1$. If $n < L$ then go to step 5.

Step 8. Rearrange the columns of T by $T = [\mathbf{t}_1 T_2]$, where $\mathbf{t}_1$ is the $q_{\max}$th column of T and T_2 is the matrix of all the remaining columns. Replace D by $D \cdot \mathrm{Diag}[I_{n-1}, T]$.

Step 9. Compute $m = m + 1$. If $m < M + 1$ then compute $G_i^{(m)} = (T_2^H T_2)^{-1} T_2^H G_i T_2$ for $1 \leq i \leq 3$, and go to Step 4.

Step 10. Same as Step 8 of the algorithm in Subsection 5.5.1.

5.5.3. The 1-D Based 2-D PESS Method with Fitness Test of 2-D Parameter Vectors

Step 1, 2. Same as that of the algorithm in Subsection 5.5.1.

Step 3. Initialize the loop; $m = 0$, $d_{\max} = -1$. Choose r_{0i} and r_{1i}, for $1 \leq i \leq 3$.

Step 4. Generate a complex random vector (g_1, g_2, g_3) where $g_i = r_i e^{j\phi_i}$ and $r_{i0} \leq r_i \leq r_{i1}$. Compute $F = g_1 G_1 + g_2 G_2 + g_3 G_3$. Compute the eigendecomposition of F by $F = T \cdot \mathrm{Diag}[\lambda_1 \cdots \lambda_M] \cdot T^{-1}$.

Step 5. Compute $\hat{A} = \hat{B}_S T$.

Step 6. Construct S_{1a}, S_{1b}, S_{2a} and S_{2b} by Eq. (5.25). Solve the linear system via the least squares method by

$$\Phi_1 = (S_{1a}^H S_{1a})^{-1} S_{1a}^H S_{1b}$$

and

$$\Phi_2 = (S_{2a}^H S_{2a})^{-1} S_{2a}^H S_{2b}.$$

Denote the set of 2-D signal zeros $\{((\Phi_1)_{ii}, (\Phi_2)_{ii})\}_{i=1}^{M}$ by $\mathcal{S}$.

Step 7. Compute $d = \sum_{i=1}^{M} \mathbf{a}_i^H B_S B_S^H \mathbf{a}_i$, where $\mathbf{a}_i = \mathbf{a}((\Phi_1)_{ii}, (\Phi_2)_{ii})$, where $\mathbf{a}(\cdot)$ is defined by Eq. (3.5). If $d > d_{\max}$, then set $\mathcal{C} = \mathcal{S}$ and $d_{\max} = d$.

Step 8. Compute $m = m + 1$. If $m < L + 1$ then go to step 4.

Step 9. Each 2-D signal zero in the set $\mathcal{C}$ gives the estimate of a 2-D wavenumber.

6. THE PAIRING-PESS METHOD

The algorithm presented in the previous section can be used as an algorithm for pairing separate 1-D parameters estimated by 1-D based 2-D methods. Assume that we have estimated 1-D signal zeros, $\{e^{j\omega_{1i_k}}\}_{k=1}^{M_1}$ and $\{e^{j\omega_{2i_k}}\}_{k=1}^{M_2}$, and we know the number M of 2-D sinusoids. Using the signal zeros estimated, construct the polynomial equations by

$$\sum_{k=0}^{P_1} t_{1k} z^k \triangleq z_1^{P_1 - M_1 - d_1} \cdot (z_1 - 1)^{d_1} \prod_{k=1}^{M_1} (z_1 - e^{j\omega_{1i_k}}), \quad (6.1.a)$$

$$\sum_{k=0}^{P_1} t_{2k} z^k \triangleq z_2^{P_2 - M_2 - d_2} \cdot (z_2 - 1)^{d_2} \prod_{k=1}^{M_2} (z_2 - e^{j\omega_{2i_k}}), \quad (6.1.b)$$

where $d_i = 0$ if $M = M_i$, or $d_i = 1$ if $M > M_i$, for $i = 1, 2$. When $M > M_i$, some of $e^{j\omega_{1i_k}}$, for $1 \leq k \leq M_i$, are multiple order zeros. The case of multiple 1-D signal zeros is accounted for by the 2-D Vandermonde system. Suppose that data matrix X in Eq. (2.2) for the case when $M = 2$ in Eq. (2.1) was available. Suppose that the estimated 1-D signal zero sets are $\{e^{j\cdot 1}\}$ and $\{e^{j\cdot 2}\}$. Here $M_1 = 1$, and $M_2 = 1$ and using these two sets of 1-D signal zeros, Eq. (6.1.a) and (6.1.b) are written as

$$z_1^{P_1 - 2} \cdot (z_1 - 1)(z_1 - e^{j\cdot 1}) = 0, \quad (6.2.a)$$

$$z_2^{P_2 - 2} \cdot (z_2 - 1)(z_2 - e^{j\cdot 2}) = 0. \quad (6.2.b)$$

When we construct the 2-D Vandermonde system, using the two 1-D polynomial equations in Eq. (6.2.a) and (6.2.b), described by Eq. (5.13), $\mathbf{a}(1, 1)$, $\mathbf{a}(1, e^{j\cdot 2})$, $\mathbf{a}(e^{j\cdot 1}, 1)$, and $\mathbf{a}(e^{j\cdot 1}, e^{j\cdot 2})$ are the eigenvectors of $D(\mathbf{g})$. Theorem 5.1 enables us to extract the subset of eigenvectors associated with the two 2-D wavenumbers. The remaining eigenvectors account for other incorrectly paired 2-D zeros and spurious 2-D zeros. When the algorithm given in Section 5 is used as a general pairing algorithm, Eq. (6.1.a) and (6.1.b) are substituted for Eq. (5.4) and (5.7). All the remaining procedure is used with no change.

6.1. The Algorithms

In the following, it is assumed that a sample data matrix X has been given as in Eq. (2.2), the number M of 2-D wavenumbers is known, and $P_1(>M)$ and $P_2(>M)$ have been chosen.

6.1.1. The Pairing-PESS Method with Distance Test of Eigenvalues

The steps are the same as those of the algorithm in Subsection 5.5.1, except that Step 2.a is changed to the following.

Step $\hat{2}$.a. Suppose that the estimates of the separate frequencies are given by the sets $\{\omega_{1i}\}_{i=1}^{M_1}$ and $\{\omega_{2i}\}_{i=1}^{M_2}$, where $M_1 < M$ and $M_2 < M$. Two univariate polynomial equations are given by Eqs. (6.1.a) and (6.1.b).

6.1.2. *The Pairing-PESS Method with Distance Test of Randomized Eigenvalues*

The steps are the same as those of the algorithm in Subsection 5.5.2, except that Step 2.a is changed to the Step $\hat{2}$.a of the algorithm in Subsection 6.1.1.

6.1.3. *The Pairing-PESS Method with Fitness Test of 2-D Parameter Vectors*

The steps are the same as those of the algorithm in Subsection 5.5.3, except that Step 2.a is changed to the Step $\hat{2}$.a of the algorithm in Subsection 6.1.1.

7. DERIVATION OF THE DIRECT 2-D PESS METHOD

In this section, it is shown that the 2-D wavenumbers, $(\omega_{1i}, \omega_{2i})$, for $1 \leq i \leq M$, are directly obtainable by using the structure of the vector $\mathbf{a}(e^{j\omega_1}, e^{j\omega_2})$ and the estimate of the range space of the matrix A defined in Eq. (3.6).

7.1. A Basic Matrix Equation

For convenience, the following convention for indexing the elements of a $(P_2 \cdot P_1) \times 1$ column vector is introduced. The $\{P_1 \cdot (j-1) + i\}$th element of the vector, where $1 \leq j \leq P_2$ and $1 \leq i \leq P_1$, will be called the (j, i)th element. Thus, the (j, i)th element of the $(P_2 \cdot P_1) \times 1$ column vector $\mathbf{a}(z_1, z_2)$, after replacing $e^{j\omega_{1i}}$ by z_1 and $e^{j\omega_{2i}}$ by z_2, is

$$a_{j,i} = z_2^{j-1} \cdot z_1^{i-1}. \tag{7.1}$$

The elements of $\mathbf{a}(z_1, z_2)$ are then shown explicitly below.

$$\mathbf{a}(z_1, z_2) = [a_{1,1}a_{1,2} \cdots a_{1,P_1} \cdots a_{P_2,1}a_{P_2,2} \cdots a_{P_2,P_1}]^t. \tag{7.2}$$

From Eq. (7.1), it follows that

$$a_{j,i} \cdot (g_2 z_2 + g_1 z_1) = g_2 a_{j+1,i} + g_1 a_{j,i+1}, \tag{7.3}$$

where $1 \leq i \leq P_1 - 1$, $1 \leq j \leq P_2 - 1$, and g_1 and g_2 are arbitrary complex numbers. Let $\mathbf{e}_{j,i}$ be the (j, i)th column of the identity matrix of order $P_2 \cdot P_1$, i.e., the $(P_2 \cdot P_1) \times 1$ column vector whose (j, i)th element is one and all the remaining elements are zero. Then, Eq. (7.1) can be expressed as

$$a_{j,i} = \mathbf{e}_{j,i}^H \mathbf{a}(z_1, z_2). \tag{7.4}$$

Eq. (7.3) can then be reformulated as

$$(g_2 z_2 + g_1 z_1) \cdot \mathbf{e}_{j,i}^H \mathbf{a}(z_1, z_2) = g_2 \mathbf{e}_{j+1,i}^H \mathbf{a}(z_1, z_2) + g_1 \mathbf{e}_{j,i+1}^H \mathbf{a}(z_1, z_2), \tag{7.5}$$

where $1 \leq i \leq P_1 - 1$ and $1 \leq j \leq P_2 - 1$. Now, let $\mathbf{e}_{j,i}$, for $1 \leq i \leq P_1 - 1$ and $1 \leq j \leq P_2 - 1$, be stacked as the matrix,

$$C = [\mathbf{e}_{1,1}\mathbf{e}_{1,2} \cdots \mathbf{e}_{1,P_1-1} \cdots \mathbf{e}_{P_2-1,1}\mathbf{e}_{P_2-1,2} \cdots \mathbf{e}_{P_2-1,P_1-1}], \tag{7.6}$$

which has $(P_1 P_2)$ rows and $(P_1 - 1)(P_2 - 1)$ columns. Stacking both sides of Eq. (7.5), it follows that

$$(g_2 z_2 + g_1 z_1) \cdot C^H \mathbf{a}(z_1, z_2) = D^H(g_1, g_2)\mathbf{a}(z_1, z_2), \tag{7.7}$$

where

$$D(g_1, g_2) = g_2 D_2 + g_1 D_1, \tag{7.8}$$

$$D_2 = [\mathbf{e}_{2,1}\mathbf{e}_{2,2} \cdots \mathbf{e}_{2,P_1-1} \cdots \mathbf{e}_{P_2,1}\mathbf{e}_{P_2,2} \cdots \mathbf{e}_{P_2,P_1-1}], \tag{7.9}$$

and

$$D_1 = [\mathbf{e}_{1,2}\mathbf{e}_{1,3} \cdots \mathbf{e}_{1,P_1} \cdots \mathbf{e}_{P_2-1,2}\mathbf{e}_{P_2-1,3} \cdots \mathbf{e}_{P_2-1,P_1}]. \tag{7.10}$$

The matrix equation (7.7) will be used to develop the new algorithm. Eq. (7.7) holds for an arbitrary 2-tuple (z_1, z_2), as is obvious from the steps in its derivation. Substituting $(e^{j\omega_{1i}}, e^{j\omega_{2i}})$ for (z_1, z_2) in Eq. (7.7) yields

$$D^H(g_1, g_2)\mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}) = \lambda_i(g_1, g_2) C^H \mathbf{a}(e^{j\omega_{1i}}, e^{j\omega_{2i}}), \tag{7.11}$$

where $1 \leq i \leq M$ and

$$\lambda_i(g_1, g_2) = g_2 e^{j\omega_{2i}} + g_1 e^{j\omega_{2i}}. \tag{7.12}$$

Through the definition of the matrix A extracted from Eq. (3.6), Eq. (7.11) leads to

$$D^H(g_1, g_2)A = C^H A \Lambda(g_1, g_2), \tag{7.13}$$

where $\Lambda(g_1, g_2)$ is a diagonal matrix whose ith diagonal element is $\lambda_i(g_1, g_2)$ and whose order is M. Eq. (7.13) can be solved for A using the property of signal (zeros) selectivity in the signal subspace, as described in the next section.

7.2. Estimation of the Matrix A

Substituting Eq. (3.19.b) for A in Eq. (7.13), the problem of solving for A is reduced to the problem of obtaining the $M \times M$ nonsingular matrix T in

$$\{D^H(g_1, g_2)B_S\} \cdot T = (C^H B_S) \cdot T \cdot \Lambda(g_1, g_2) \quad (7.14)$$

Using Eqs. (3.6) and (7.6), it can be shown that

$$C^H A = [\mathbf{a}'(e^{j\omega_{11}}, e^{j\omega_{21}}) \cdots \mathbf{a}'(e^{j\omega_{1M}}, e^{j\omega_{2M}})], \quad (7.15)$$

where

$$\mathbf{a}'(e^{j\omega_{1i}}, e^{j\omega_{21}}) = \mathbf{a}_2'(e^{j\omega_{2i}}) \otimes \mathbf{a}_1'(e^{j\omega_{1i}}),$$
$$\mathbf{a}_2'(e^{j\omega_{2i}}) = [1 e^{j\omega_{2i}} \cdots e^{j\omega_{2i}(P_2-2)}],$$

and

$$\mathbf{a}_1'(e^{j\omega_{1i}}) = [1 e^{j\omega_{1i}} \cdots e^{j\omega_{1i}(P_1-1)}].$$

From the discussion in Section 3.3 and [4], it is known that $C^H A$ is of full rank if and only if

$$(P_1 - 1) \cdot (P_2 - 1) \geq M. \quad (7.16)$$

Condition (7.16) is always satisfied since it was assumed that $P_1 > M$ and $P_2 > M$. Since A and B_S span the same signal subspace, $C^H B_S$ is also of full rank. Thus, Eq. (7.14) implies that

$$F(g_1, g_2) = T \cdot \Lambda(g_1, g_2) \cdot T^{-1}, \quad (7.17)$$

where

$$F(g_1, g_2) = (B_S^H C C^H B_S)^{-1} B_S^H C D^H(g_1, g_2) B_S. \quad (7.18)$$

The $M \times M$ non-singular matrix T and the diagonal matrix $\Lambda(g_1, g_2)$ can be obtained by eigendecomposing the matrix $F(g_1, g_2)$. Eq. (7.17) implies that the matrix F is diagonalizable (non-defective) [20, p. 338].

LEMMA 7.1. *With a distinct eigenvalue is associated one eigenvector which is unique up to scalar multiplication.*

Proof. $F \in \mathbf{C}^{M \times M}$ can be Schur-decomposed as $Q^H F Q = T$, where $Q \in \mathbf{C}^{M \times M}$ is a unitary matrix and $T \in \mathbf{C}^{M \times M}$ is a upper triangular matrix [20, pp. 335]. Since $\det(F - \lambda I) = \det(QTQ^H - \lambda I) = \det(T - \lambda I)$, the diagonal elements λ_i, for $1 \leq i \leq M$, are the eigenvalues of F. Consider an eigenvalue λ_i. Its associated eigenvector x_i satisfies $(F - \lambda_i I)x_i = 0$ or $(T - \lambda_i I)Q^H x_i = 0$. If λ_i is distinct, $T - \lambda_i I$ is of rank $M - 1$. Thus, $Q^H x_i$ is unique up to scalar multiplication, and so is x_i. ■

When the eigenvalues of the matrix $F(g_1, g_2)$ are distinct ($F(g_1, g_2)$ is nonsingular if and only if any eigenvalue is not zero.), the columns of the eigenvector matrix T are unique up to scalar multiplication, and so the estimate of A is given by

$$A = B_S T \Delta, \quad (7.19)$$

where Δ is a diagonal matrix whose elements are determined to scale each column of $B_S T$ such that the first element is 1. The estimate of 2-D wavenumbers are computed from Eq. (7.19) using Eqs. (5.23), (5.24), and (5.25).

7.3. The Algorithms

In the following, it is assumed that a sample data matrix X has been given as in Eq. (2.2), the number M of 2-D wavenumbers is known, and $P_1(>M)$ and $P_2(>M)$ have been chosen.

7.3.1. The Direct 2-D PESS Method with Distance Test of Eigenvalues

The steps are the same as those of the algorithm in Subsection 5.5.1, except that g_3 and G_3 vanish in all the steps and the Steps 2.a and 2.b are changed to the following.

Step 2. D_2 and D_1 are constructed by Eqs. (7.9) and (7.10), and the following 2 matrices are computed.

$$G_1 = (\hat{B}_S^H C C^H B_S)^{-1} B_S^H C D_1^H B_S$$

and

$$G_2 = (\hat{B}_S^H C C^H B_S)^{-1} B_S^H C D_2^H B_S.$$

7.3.2. The Direct 2-D PESS Method with Distance Test of Randomized Eigenvalues

The steps are the same as those of the algorithm in Subsection 5.5.2, except that g_3 and G_3 vanish in all the steps and the Steps 2.a and 2.b are changed to the step 2 of the algorithm in Subsection 7.3.1.

7.3.3. The Direct 2-D PESS Method with Fitness Test of 2-D Parameter Vectors

The steps are the same as those of the algorithm in Subsection 5.5.3, except that g_3 and G_3 vanish in all the steps and the Steps 2.a and 2.b are changed to the Step 2 of the algorithm in Subsection 7.3.1.

8. SIMULATION RESULTS

In this section, examples are presented to illustrate the PESS methods developed above, and Monte Carlo simulation results are provided to show the performance of the methods. In the simulation, the PESS method and the MEMP method [8] were applied to the same data set for accuracy comparison.

Each 20×20 sample data matrix was generated by the equation

$$x(m, n) = \sum_{k=1}^{3} e^{j2\pi(f_{2k}m + f_{2k}n)} + v(m, n), \quad \text{for } 0 \leq m, n \leq 19, \quad (8.1)$$

where $v(m, n)$ is the complex white noise.

8.1. Examples

The 3 wavenumber parameters were selected as

$$(f_{11}, f_{21}) = (0.26, 0.24) \quad (8.2)$$

$$(f_{12}, f_{22}) = (0.24, 0.24) \quad (8.3)$$

$$(f_{13}, f_{23}) = (0.24, 0.26). \quad (8.4)$$

The 3 algorithms processed a sample data matrix X which was generated with SNR of 20 dB. SNR is defined by

$$SNR = 10 \log_{10}(1/\sigma^2), \quad (8.5)$$

where σ^2 is the noise power. The 4×4 array was selected, which means that $P_1 = P_2 = 4$. The number L of iteration was set to 2 and $r_{0i} = 1$ and $r_{1i} = 10$, for $1 \leq i \leq 3$, were selected in Sections 8.1.2 and 8.1.3.

8.1.1. The 1-D Based 2-D PESS Method with Distance Test of Eigenvalues

Step 1. The list of data matrix X, $\hat{R}$, $\hat{R}_1$, $\hat{R}_2$, $\hat{B}$, $\hat{\Delta}$, $\hat{B}_1$, $\hat{\Delta}_1$, $\hat{B}_2$, and $\hat{\Delta}_2$ were obtained and is available from the authors as an Appendix to this manuscript.

Step 2.a. Two univariate polynomial equations in z_i, for $i = 1, 2$, were computed as

$$0.0 + (-0.0204075 + j0.1130556)z_1 + (-0.2097420 + j0.0769447)z_1^2$$
$$+(-0.2097420 - j0.0769447)z_1^3 + (-0.0204075 - j0.1130556)z_1^4 = 0.0$$

and

$$0.0 + (0.3638633 - j0.4630003)z_2 + (0.6093172 + j0.2134817)z_2^2$$
$$+(0.6093172 - j0.2134817)z_2^3 + (0.3638633 + j0.4630003)z_2^4 = 0.0.$$

Step 2.b. The initial G_1, G_2 and G_3 were computed as

$$G1 = \begin{pmatrix} 0.0236260 + j0.9975650 & 0.0321896 - j0.0225824 & -0.0000030 - j0.0015337 \\ 0.0686103 + j0.0446103 & -0.0180299 + j1.0018485 & -0.0025797 + j0.0040438 \\ 0.0155744 + j0.6278006 & -0.3369688 - j0.5409832 & 0.0179116 + j0.9976731 \end{pmatrix},$$

$$G2 = \begin{pmatrix} 0.0177350 + j0.9998431 & -0.0400412 + j0.0150251 & 0.0001062 - j0.0011839 \\ -0.0760086 - j0.0483654 & -0.0297496 + j0.9890730 & 0.0025833 - j0.0035922 \\ 0.0096372 + j0.6282084 & 0.3278028 + j0.5331685 & 0.0181662 + j0.9980672 \end{pmatrix},$$

and

$$G3 = \begin{pmatrix} -1.0044686 + j0.0348798 & 0.0043723 - j0.0093423 & 0.0027056 - j0.0000879 \\ 0.0038265 - j0.0068570 & -0.9936933 - j0.0474929 & -0.0003874 - j0.0000307 \\ -1.2538954 + j0.1007968 & 0.0082100 - j0.0070122 & -0.9974769 + j0.0359509 \end{pmatrix}.$$

Step 3. $\mathbf{g} = [0\ 0\ 1]$, $\mathbf{g}' = [1\ -0.5\ 0]$ and $\delta = 0.07$ were selected.

Step 4. $F = g_1G_1 + g_2G_2 + g_3G_3$ was computed as

$$F = \begin{pmatrix} -1.0044686 + j0.0348798 & 0.0043723 - j0.0093423 & 0.0027056 - j0.0000879 \\ 0.0038265 - j0.0068570 & -0.9936933 - j0.0474929 & -0.0003874 - j0.0000307 \\ -1.2538954 + j0.1007968 & 0.0082100 - j0.0070122 & -0.9974769 + j0.0359509 \end{pmatrix}.$$

The eigenvalue and the eigenvector matrices of F were obtained as

$$\Delta = Diag[-1.0048345 - j0.0229922 - 0.9980068 + j0.0940336 - 0.9927975 - j0.0477037]$$

and

$$T = \begin{pmatrix} 0.0021124 + j0.0471462 & 0.0041338 - j0.0459490 & -0.0075787 + j0.0682349 \\ 0.0054370 - j0.0001844 & -0.0015099 + j0.0048979 & -0.0158513 + j0.3115187 \\ 1.0000000 + j0.0000000 & 1.0000000 + j0.0000000 & 1.0000000 + j0.0000000 \end{pmatrix}.$$

Step 5. The distances between the eigenvalues were computed as in the distance matrix

$$\begin{pmatrix} 0.0000000 & 0.1172248 & 0.0274872 \\ 0.1172248 & 0.0000000 & 0.1418330 \\ 0.0274872 & 0.1418330 & 0.0000000 \end{pmatrix}.$$

The clustering procedure gave cluster 1: {2} and cluster 2: {1, 3}.

Step 6. Consider cluster 1. Keep the second column of T.

Step 7. Consider cluster 2. The following F' was computed with $\mathbf{g}' = [1\ -0.5\ 0]$.

$$F' = \begin{pmatrix} 0.0430450 + j0.5375658 & 0.0366425 + j0.0193354 \\ 0.2040451 - j0.1839002 & -0.0454788 + j0.4761306 \end{pmatrix}.$$

The eigendecomposition $F' = T'\Delta'(T')^{-1}$ was computed as

$$\Delta' = Diag[-0.1109810 + j0.50718540.1085473 + j0.5065110]$$

and

$$T' = \begin{pmatrix} -0.2528226 - j0.0756661 & 0.3424790 + j0.4575574 \\ 1.0000000 + j0.0000000 & 1.0000000 + j0.0000000 \end{pmatrix}.$$

Step 8. The desired eigenvector matrix was obtained as

$$\hat{T} = \begin{pmatrix} -0.0317045 + j0.0670475 & 0.0041338 - j0.0459490 & 0.0091916 + j0.0304683 \\ -0.1425295 + j0.0992513 & -0.0015099 + j0.0048979 & 0.0330160 - j0.0777440 \\ 1.3424790 + j0.4575574 & 1.0000000 + j0.0000000 & 0.7471774 - j0.0756661 \end{pmatrix}.$$

The estimate of A was computed as

$$\hat{A} = \begin{pmatrix} 0.3513742 + j0.0823592 & 0.2483042 + j0.0218955 & 0.1825521 - j0.0333121 \\ -0.0606784 + j0.3559272 & -0.0092024 + j0.2492167 & 0.0184874 + j0.1869526 \\ -0.3587124 - j0.0390216 & -0.2493843 + j0.0033865 & -0.1901246 + j0.0043823 \\ 0.0153205 - j0.3609187 & -0.0156432 - j0.2488627 & 0.0108940 - j0.1925282 \\ -0.1142487 + j0.3388163 & -0.0107822 + j0.2498093 & 0.0452373 + j0.1799749 \\ -0.3464285 - j0.0922990 & -0.2499995 + j0.0016945 & -0.1853295 + j0.0305067 \\ 0.0718142 - j0.3515703 & -0.0142784 - j0.2497213 & -0.0168328 - j0.1891642 \\ 0.3557070 + j0.0490836 & 0.2486869 - j0.0266547 & 0.1924566 - j0.0018425 \\ -0.3256860 - j0.1427773 & -0.2503781 + j0.0002922 & -0.1771500 + j0.0557414 \\ 0.1216560 - j0.3349905 & -0.0127796 - j0.2502443 & -0.0414145 - j0.1831561 \\ 0.3420828 + j0.1018271 & 0.2493742 - j0.0253760 & 0.1880830 - j0.0280891 \\ -0.0794458 + j0.3480729 & 0.0377740 + j0.2480116 & 0.0131399 + j0.1921169 \\ 0.1726683 - j0.3085556 & -0.0111399 - j0.2505019 & -0.0674049 - j0.1734022 \\ 0.3199914 + j0.1519383 & 0.2498546 - j0.0235370 & 0.1802734 - j0.0533663 \\ -0.1325859 + j0.3296704 & 0.0361808 + j0.2485306 & 0.0400952 + j0.1859581 \\ -0.3379245 - j0.1122691 & -0.2464574 + j0.0488280 & -0.1908892 + j0.0256973 \end{pmatrix}.$$

From $\hat{A}$, the following two matrices were obtained.

$$\Phi_1 = \begin{pmatrix} 0.0662453 + j0.9975544 & 0.0008960 - j0.0005833 & -0.0012975 - j0.0003834 \\ 0.0022932 + j0.0024737 & 0.0493609 + j0.9988683 & -0.0021381 + j0.0109805 \\ -0.0099251 - j0.0038398 & -0.0010368 + j0.0006666 & -0.0728221 + j0.9969979 \end{pmatrix}.$$

$$\Phi_2 = \begin{pmatrix} -0.0835589 + j0.9959845 & -0.0023614 + j0.0024702 & 0.0025242 - j0.0003328 \\ -0.0182159 - j0.0148995 & 0.0472167 + j0.9988186 & -0.0016406 - j0.0004648 \\ 0.0119660 + j0.0059019 & 0.0012700 - j0.0000055 & 0.0606975 + j0.9979546 \end{pmatrix}.$$

The desired 2-D zeros were extracted from the above two matrices as

$$(0.9997516 \cdot exp(j2\pi 0.2394464),\ 0.9994834 \cdot exp(j2\pi 0.2633212)),$$
$$(1.0000872 \cdot exp(j2\pi 0.2421415),\ 0.9999340 \cdot exp(j2\pi 0.2424819)),$$
$$(0.9996539 \cdot exp(j2\pi 0.2616043),\ 0.9997987 \cdot exp(j2\pi 0.2403318)).$$

8.1.2. The Pairing-PESS Method with Distance Test of Randomized Eigenvalues

It was assumed that two sets of 1-D wavenumber parameters were estimated as

$$\{0.2595620,\ 0.2396049\},$$
$$\{0.2567192,\ 0.2431879\}.$$

Two univariate polynomial equations in z_i, for $i = 1, 2$, were obtained as

$$0.0 + (0.9999863 - j0.0052345)z_1 + (-0.9947621 + j2.0012980)z_1^2$$
$$+(-1.0052242 - j1.9960635)z_1^3 + 1.0z_1^4 = 0$$

and

$$0.0 + (0.9999998 - j0.0005840)z_2 + (-0.9994164 + j1.9987771)z_2^2$$
$$+(-1.0005834 - j1.9981931)z_2^3 + 1.0z_2^4 = 0.$$

Iteration 1. The vector **g** was randomly generated as

$$\mathbf{g} = [8.2824010 + j2.7776125\ \ 1.7048545 - j4.0095459\ \ 0.3119416 - j0.9788936]^t.$$

The maxmin distance was computed as 1.0368955. Another **g** was randomly generated as

$$\mathbf{g} = [-0.1596835 + j2.1919652\ \ 6.8230877 + j7.0010181\ \ -6.1613831 - j8.6783894]^t.$$

The maxmin distance was computed as 1.0061741. The first trial was determined to be better. The matrix D was obtained as

$$\begin{pmatrix} 0.0068896 + j0.0257567 & 0.0072275 + j0.0205755 & 0.0029841 - j0.0483550 \\ 0.0476577 - j0.0772910 & -0.0524261 + j0.0925424 & -0.0107985 + j0.0038552 \\ 1.0000000 + j0.0000000 & 1.0000000 + j0.0000000 & 1.0000000 + j0.0000000 \end{pmatrix}.$$

Iteration 2. The vector **g** was randomly generated as

$$\mathbf{g} = [5.4793580 + j6.2770135\ \ -1.5739585 + j0.5342874\ \ 2.0556196 - j3.2626202]^t.$$

The maxmin distance was computed as 1.6158669. Another **g** was randomly generated as

$$\mathbf{g} = [5.8321917 + j7.7009918\ \ -0.2320668 + j3.1577547\ \ g3 = -4.3553791 - j9.0153360]^t.$$

The maxmin distance was computed as 1.8360912. The matrix D was updated as

$$\begin{pmatrix} 0.0068896 + j0.0257567 & 0.0028700 - j0.0482168 & -0.0066433 - j0.0632031 \\ 0.0476577 - j0.0772910 & -0.0116724 + j0.0038789 & 0.0128859 - j0.0791381 \\ 1.0000000 + j0.0000000 & 1.0042443 + j0.0070381 & 0.2113152 + j0.1908670 \end{pmatrix}.$$

The desired 2-D zeros were extracted from the above two matrices as

$$(1.0003608 \cdot exp(j2\pi 0.2615938),\ 1.0001100 \cdot exp(j2\pi 0.2402652)),$$
$$(1.0002271 \cdot exp(j2\pi 0.2419204),\ 0.9929687 \cdot exp(j2\pi 0.2431732)),$$
$$(0.9988808 \cdot exp(j2\pi 0.2396621),\ 1.0055987 \cdot exp(j2\pi 0.2625731)).$$

8.1.3. The Direct 2-D PESS Method with Fitness Test of 2-D Parameter Vectors

Trial 1. The following were generated randomly.

$$g1 = 8.2824010 + j2.7776125 \quad \text{and} \quad g2 = 1.7048545 - j4.0095459.$$

The eigenvector matrix was computed as

$$T =$$

$$\begin{pmatrix} -0.0022579 + j0.0378996 & -0.0061035 + j0.032090 & -0.0024479 - j0.0504817 \\ 0.0563264 - j0.0876339 & -0.0741348 + j0.087846 & 0.0008896 - j0.0042041 \\ 1.0000000 + j0.0000000 & 1.0000000 + j0.000000 & 1.0000000 + j0.0000000 \end{pmatrix}.$$

With this estimate of T, the signal zeros were computed as

$$(-0.0729085 + j0.9971211,\ 0.0611014 + j0.9980343)$$
$$(0.0666534 + j0.9975640,\ -0.0843249 + j0.9953308)$$
$$(0.0490392 + j0.9987356,\ 0.0475788 + j0.9993925).$$

and the criterion value was evaluated as 47.9998695.

Trial 2. The following were generated randomly.

$$g(1) = -0.1596835 + j2.1919652 \quad \text{and} \quad g(2) = 6.8230877 + j7.0010181.$$

The eigenvector matrix was computed as

$$T' =$$

$$\begin{pmatrix} -0.0075360 + j0.0414075 & -0.0001785 - j0.033024 & -0.0139303 + j0.0298457 \\ -0.0666335 + j0.1067764 & 0.0090331 - j0.015174 & 0.0623083 - j0.0678467 \\ 1.0000000 + j0.0000000 & 1.0000000 + j0.000000 & 1.0000000 + j0.0000000 \end{pmatrix}.$$

With this estimate of T, the signal zeros were computed as

$$(0.0664314 + j0.9976989,\ -0.0845125 + j0.9959812)$$
$$(0.0513499 + j0.9941209,\ 0.0470659 + j0.9994206)$$
$$(-0.0749972 + j1.0016009,\ 0.0618018 + j0.9973559).$$

and the criterion value was evaluated as 47.9998661.

The result of the first trial was determined to be best. The desired 2-D zeros were extracted from the above two matrices as

$$(0.9997830 \cdot exp(j2\pi 0.2616166),\ 0.9999029 \cdot exp(j2\pi 0.2402684)),$$
$$(0.9997883 \cdot exp(j2\pi 0.2393817),\ 0.9988964 \cdot exp(j2\pi 0.2634516)),$$
$$(0.9999388 \cdot exp(j2\pi 0.2421916),\ 1.0005244 \cdot exp(j2\pi 0.2424287)).$$

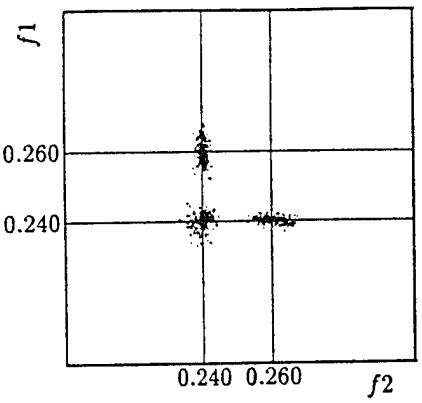


FIG. 8.1. The 1-D based 2-D PESS; 20 dB; 200 runs; $P_1 = P_2 = 5$.

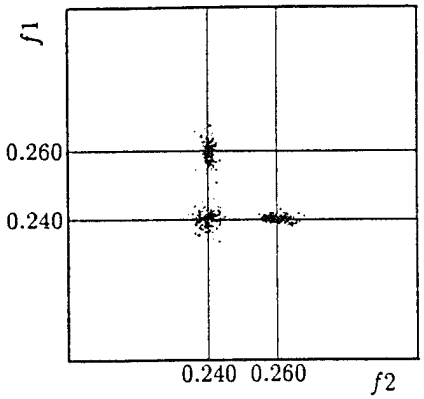


FIG. 8.2. The pairing-PESS; 20 dB; 200 runs; $P_1 = P_2 = 4$.

8.2. Monte Carlo Run

The simulation results for the 1-D based 2-D PESS method, the Pairing-PESS method, and the Direct 2-D PESS method are presented. The procedure for avoiding the problem of repeated eigenvalues was selected to provide the fitness test of 2-D parameter vectors. Comparison of the PESS methods with the MEMP method was also made. Two hundred sample data matrices were generated and the three PESS methods and the MEMP method were applied to estimate the 2-D wavenumbers. The comparisons between the four methods were made for the wavenumber parameters in Eqs. (8.2), (8.3), and (8.4). (The results are shown in Figs. 8.1 to 8.4.)

Subsequently, the Direct 2-D PESS method and the MEMP method were compared for wavenumber parameters specified by

$$(f_{11}, f_{21}) = (0.26, 0.24) \tag{8.6}$$
$$(f_{12}, f_{22}) = (0.24, 0.24) \tag{8.7}$$
$$(f_{13}, f_{23}) = (0.25, 0.26). \tag{8.8}$$

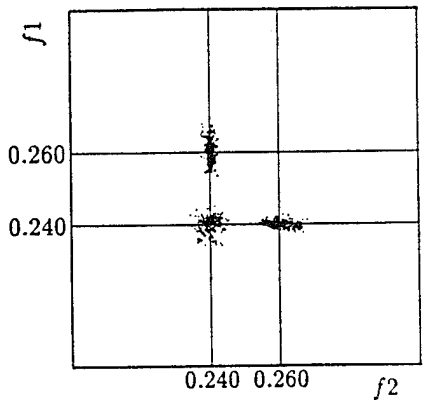


FIG. 8.3. The direct 2-D PESS; 200 runs; $P_1 = P_2 = 4$.

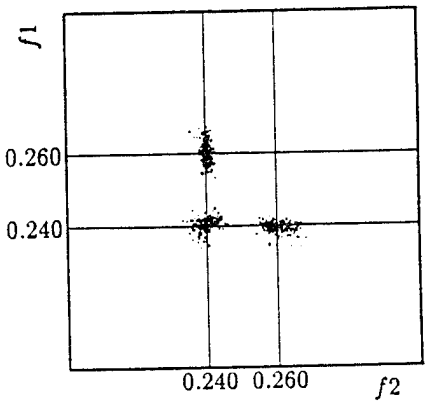


FIG. 8.4. The MEMP; 20 dB; 200 runs; $P_1 = P_2 = 4$.

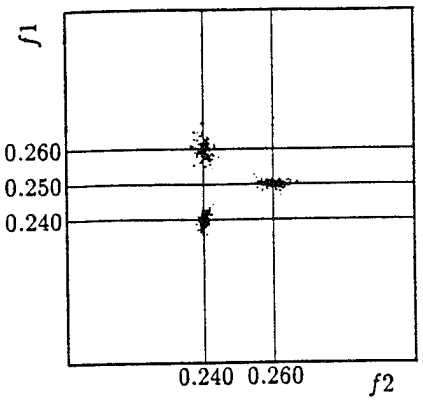


FIG. 8.6. The direct 2-D PESS; 20 dB; 200 runs; $P_1 = P_2 = 5$.

For the preceding set the parameters P_1 and P_2 were each selected to be 5 to improve resolution. The results for this case are shown in Figs. 8.5 and 8.6.

In the simulations, $r_{0i} = 1$ and $r_{1i} = 10$, for $1 \leq i \leq 3$, were selected. The number L of iterations was chosen to be 2 for the PESS algorithms. In the simulation of the Pairing-PESS method, it was assumed that the estimates of wavenumbers, obtained by other method, include an estimation error with a standard deviation of 0.003. In the pairing procedure by the MEMP method, the criterion of Eq. (5.27) was used as recommended by Hua [8].

Figures 8.1 to 8.4 show that the PESS methods worked as well as the MEMP method. The Direct 2-D PESS method outperformed the MEMP method for the 2-D wavenumber parameters in Eqs. (8.6) to (8.8), as shown in Figs. 8.5 and 8.6.

9. CONCLUSIONS

The methods proposed here for estimation of 2-D wavenumbers of a finite number of sinusoids in noise are computationally efficient and accurate. The knowledge of the number of sinusoids or its estimate is needed but this constraint can be relaxed as in the 1-D case [3]. The procedure, in principle, is generalizable to dimensions higher than two. Though the computations are likely to increase substantially with increase in the number of dimensions, other 1-D based methods which use a separate pairing procedure as well as other direct methods are usually more computation-intensive.

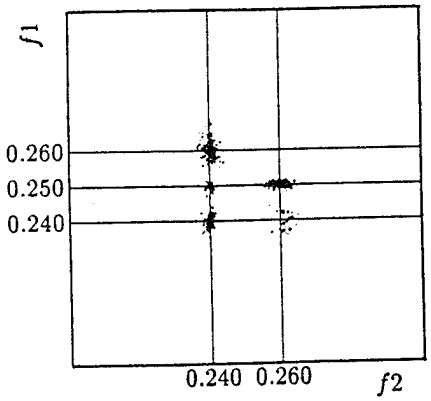


FIG. 8.5. The MEMP; 20 dB; 200 runs; $P_1 = P_2 = 5$.

REFERENCES

1. Bose, N. K. *Applied Multidimensional Systems Theory.* Van Nostrand Reinhold, New York, 1982.
2. Bose, N. K., and Basu, S. Tests for polynomial zeros on a polydisc distinguished boundary. *IEEE Trans. Circuits Syst.* **CAS25** 9 (Sep. 1978), 684–693.
3. Chun, J., and Bose, N. K. A novel subspace-based approach to parameter estimation. *Digital Signal Process.* 4 (Jan. 1994), 40–48.
4. Cabrera, S. D., and Bose, N. K. Prony methods for two-dimensional complex exponential signal modeling, *Applied Control* (Spyros G. Tzafestas, Ed.), Chap. 15, pp. 401–411. Dekker. New York, 1993.
5. Kung, S. Y., Arun, K. S., and Bhaskar Rao, D. V. State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem, *J. Opt. Soc. Am.* **73** 12 (Dec. 1983), 1799–1811.
6. Shaw, A. K., and Kumaresan, R. Frequency-wavenumber estimation by structured matrix approximation, *Proc. IEEE ASSP 3rd Workshop Spectral Estimation Modeling,* Boston. MA, 1986, pp. 81–84.
7. Zoltowski, M. D., and Stavrinides, D. Sensor array signal processing via a procrustes rotations based eigenanalysis of the ESPRIT data pencil, *IEEE Trans. Acoust. Speech Signal Process.* **37** 6 (June 1989), 832–861.
8. Hua, Y. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil, *IEEE Trans. Signal Process.* 40 9 (Sep. 1992), 2267–2280.
9. Swindlehurst, A., and Kailath, T. Azimuth/elevation direction finding using regular array geometries. *IEEE Trans. Aerosp. Elect. Syst.* **29** 1 (Jan. 1993), 145–156.

10. Chen, Y.-H., and Chen, C.-H. Direction-of-arrival and frequency estimations for narrowband sources using two single rotation invariance algorithms with the marked subspace. *IEE Proc. Part F Radar Signal Process.* **139** 4 (Aug. 1992), 297–300.
11. Kumaresan, R., and Tufts, D. W. A two-dimensional technique for frequency-wavenumber estimation. *Proc. IEEE* **69** 11 (Nov. 1981), 1515–1517.
12. Johnson, D. H. The application of spectral estimation methods to bearing estimation problems. *Proc. IEEE* **70** 9 (Sep. 1982), 1018–1028.
13. McClellan, J. H. Multidimensional spectral estimation. *Proc. IEEE* **70** 9 (Sep. 1982), 1029–1039.
14. Chen, Y.-M., Lee, J.-H., and Yeh, C.-C. Two-dimensional angle-of-arrival estimation for uniform planar arrays with sensor position errors. *IEE Proc. Part F: Radar Signal Process.* **140** 1 (Feb. 1993), 37–41.
15. van der Veen, A. J., Ober, P. B., and Deprettere, E. F. Azimuth and elevation computation in high resolution DOA estimation. *IEEE Trans. Signal Process.* **40** 7 (July 1992), 1828–1832.
16. Johnson, R. L., and Miner, G. E. An operational system implementation of the ESPRIT DF algorithm. *IEEE Trans. Aerosp. Elect. Syst.* **27** 1 (Jan. 1991), 159–166.
17. Barbieri, M. M., and Barone, P. A two-dimensional Prony's method for spectral estimation. *IEEE Trans. Signal Process.* **40** 11 (Nov. 1992), 2747–2756.
18. Kumaresan, R., and Tufts, D. W. Estimating the angle of arrival of multiple plane wave. *IEEE Trans. Aerosp. Electon. Syst.* **AES19** 1 (Jan. 1983), 134–139.
19. Cullen, C. G., *Matrices and Linear Transformations.* Addison–Wesley, Philippines, 1972.
20. Golub, G. H., and Van Loan, C. F. *Matrix Computations* Johns Hopkins Univ. Press, Baltimore, 1991.

JONGTAE CHUN was born on June 13, 1962. He received his B.S. and M.S. in instrumentation and control at Seoul National University in 1983 and 1985, respectively. Between 1985 and 1987, he worked as a researcher on development of data communication equipment at Korea Data Communication Corporation, Seoul. From 1987 to 1991, he was a full-time consultant and instructor on design of automated factories at Korea Productivity Center, Seoul. Since September 1991, he has been a graduate student in electrical engineering at The Pennsylvania State University, University Park, where he is a research assistant in the Spatial and Temporal Signal Processing Center, working toward his Ph.D. degree. His current research interests are in sensor signal and image processing, multidimensional signal processing, and neural networks.

N. K. BOSE was born on August 19, 1940. He received the B.Tech (Hons.) degree, the M.S. degree, and the Ph.D. degree from the Indian Institute of Technology, Kharagpur, Cornell University, Ithaca, NY, and Syracuse University, Syracuse, NY, respectively. He is currently the HRB-Systems Professor of Electrical Engineering and the Director of The Spatial and Temporal Signal Processing Center at The Pennsylvania State University, University Park. Dr. Bose is the author of *Applied Multidimensional Systems Theory* (Van Nostrand, 1982), *Digital Filters* (North-Holland, 1985, Krieger Publishing, 1993), and the editor as well as the main author of *Multidimensional Systems: Progress, Directions and Open Problems* (Reidel, 1985). He was the guest editor of the *Proceedings of the IEEE* special issue on Multidimensional Systems in June 1977, the *Proceedings of the IEEE* special issue on Multidimensional Signal Processing in April 1990, and the editor of the IEEE Press Selected Reprint Series book *Multidimensional Systems: Theory and Applications* (1979). He also served as guest editor of a Special Issue on Aspects of Spatial and Temporal Signal Processing for the journal *Circuits, Systems, and Signal Processing.* Dr. Bose was a guest co-editor for a special issue on Linear Algebra in Electrical Engineering which was published by *Linear Algebra and Its Applications* in January 1988. Dr. Bose is a co-editor of the book *Signal Processing and Its Applications,* published by North-Holland in 1993. He has been the editor-in-chief of the *Journal on Multidimensional Systems and Signal Processing* since 1990. He has been a Fellow of IEEE since 1981. His current research interests are in multidimensional signal processing, multitarget tracking, and neural networks.

PERFORMANCE ANALYSIS OF THE TLS ALGORITHM FOR IMAGE RECONSTRUCTION FROM A SEQUENCE OF UNDERSAMPLED NOISY AND BLURRED FRAMES

N. K. Bose, H. C. Kim, and B. Zhou

Department of Electrical Engineering
The Spatial and Temporal Signal Processing Center
The Pennsylvania State University
University Park, PA 16802
Tel: (814) 865-3912, Fax: (814) 865-7065
Email: NKB@STSPNKB.PSU.EDU

ABSTRACT

A theoretical analysis is proposed to evaluate the robustness of the total least square (TLS) algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames in the wavenumber domain. Since the data samples in the wavenumber domain are complex-valued, the results from the TLS theory developed for the case of real-valued data are adapted to the situation at hand. It is shown that the image quality can be improved as more frames become available. In the case of blurred frames, higher resolution images may be reconstructed using the TLS algorithm with post-deblurring. Finally, computer simulation results are provided to demonstrate the robustness of the TLS algorithm for image reconstruction.

1. INTRODUCTION

Recently, considerable attention has been devoted to image sequence processing. Many applications of image sequence processing may be found in (both military and industrial) surveillance, medical imaging, and agriculture. One of the efforts in image sequence processing is directed at the improving of resolution of the reconstructed image from a sequence of snapshots taken over a fixed object. For example, LANDSAT follows a circular, near-polar orbit at a nominal altitude of $700km$ and it flies over the earth surface between latitudes 82° north and 82° south once every 16 days. Thus, multiple images of the same area are available. The spatial resolution of an image is often determined by imaging sensors. In a CCD camera, the image resolution is determined by the size of its photo-detector. When an ensemble of several shifted images are available, we may reconstruct an image with higher resolution which is equivalent to an effective increase of the sampling rate by interpolation.

The problem of reconstructing a high resolution interpolated image from a sequence of undersampled noisy and blurred frames has been tackled by procedures developed both in the space domain and the wavenumber domain. In the space domain, several methods were reported [1, 2, 3]. Peleg, *et al.*, [1] proposed an interesting scheme. Starting with an initial guess for a high resolution image, they simulated the imaging process to compute low resolution images. The high resolution image is iteratively improved to achieve simulated low resolution images closest to the observed ones. Tekalp, *et al.*, [2] proposed a scheme based on a prior POCS (projection onto convex sets) method. Their procedure was successfully applied to blurred noisy images. However, the recursive scheme was not explicit. Ur and Gross [3] proposed a nonuniform interpolation scheme based on a generalized sampling theorem to obtain an improved resolution image from an ensemble of subpixel level shifted low resolution images of the same scene. They did not take into account the presence of noise.

Using sequential estimation theory in the wavenumber domain, an efficient method was developed [4] for recursively updating to provide satisfactory reconstruction in the blur-free case provided the displacements of the frames with respect to a reference frame were either known or estimated. It was observed that the performance deteriorated when the blur produced zeros in the wavenumber domain and theoretical justification

This research was partially supported by SDIO/IST and managed by the Office of Naval Research under Contract N00014-92-J-1755

0-8186-6950-0/94 $4.00 © 1994 IEEE

for this can be provided. The problem of reconstruction in the wavenumber domain with errors present both in the observation and data was tackled by the method of Total Least Squares (TLS) [5]. A recursive updating scheme was developed which performed satisfactorily in the blur-free case. Very recently, Kim [6] proposed that the blurred and noisy case be tackled by applying the method in [5] with post-deblurring.

In this paper, we expand the application of the procedures in [5] and [6] to a wide variety of blur and noise and then work towards a theoretical analysis for robustness evaluation of the TLS method in the wavenumber domain where the data is complex-valued. For this to be possible, the results from the TLS theory for the case of real-valued data are adapted to the situation at hand. In Section 2, a brief review of high resolution image reconstruction by the TLS algorithm is provided. Then, the robustness of the TLS algorithm for image reconstruction is evaluated through variance analysis in Section 3. In Section 4, some computer simulation results are supplied to demonstrate the robustness of the TLS algorithm for image reconstruction. Finally, concluding remarks are made in Section 5.

2. HIGH RESOLUTION IMAGE RECONSTRUCTION BY THE TLS ALGORITHM

Suppose that the Fourier transform, $F^c(u,v)$, of the original continuous image $f(x,y)$ is approximated by the bandlimited constraint

$$|F^c(u,v)| = 0, \quad \text{for } |u| > L_x w_x \text{ and } |v| > L_y w_y, \quad (1)$$

where L_x, L_y are some finite integers, $w_x = 2\pi/T_x$, $w_y = 2\pi/T_y$, and T_x, T_y are, respectively, the sampling periods along the x and y axes. At least $4L_xL_y$ undersampled frames, $f_i(i_x,i_y)$, $i = 1,2,\ldots,4L_xL_y$, are needed for the initial reconstruction of $f(x,y)$, where i_x and i_y are indices of natural numbers. To accommodate explicitly the non-unity sampling periods and shifts, the ith frame is expressed as

$$f_i(i_x,i_y) = f(i_xT_x + \delta_{xi}, i_yT_y + \delta_{yi}), \quad (2)$$

where $i_x = 0,1,\ldots,M-1$, $i_y = 0,1,\ldots,N-1$, and δ_{xi}, δ_{yi} are, respectively, the estimated shifts along the x and y axes. It is assumed that differences between the true shifts and the estimated ones along the x and y axes are, respectively, $\Delta\delta_{xi}$ and $\Delta\delta_{yi}$. If there are k frames available, the multiframe image restoration model is [5]

$$\boldsymbol{Z} = [\boldsymbol{\Phi} + \boldsymbol{E}]\boldsymbol{F} + \boldsymbol{N}, \quad (3)$$

where

$$\begin{aligned}
\boldsymbol{Z} &= [Z_1\ Z_2\ \ldots\ Z_k]^t, \quad \boldsymbol{\Phi} = [\boldsymbol{Y}_1\ \boldsymbol{Y}_2\ \ldots\ \boldsymbol{Y}_k]^t,\\
\boldsymbol{Y}_i &= [\phi_{i1}\ \phi_{i2}\ \ldots\ \phi_{ip}]^t,\\
\phi_{i\tau} &= \frac{1}{T_xT_y} exp\left(j2\pi\left(\delta_{xi}\left(\frac{m}{MT_x}+\frac{l_x}{T_x}\right)\right.\right.\\
&\quad \left.\left.+\delta_{yi}\left(\frac{n}{NT_y}+\frac{l_y}{T_y}\right)\right)\right),\\
l_x &= (\tau-1) mod(2L_x) - L_x,\\
l_y &= \lfloor(\tau-1)/(2L_x)\rfloor - L_y,\\
\Delta\boldsymbol{Y}_i &= [\varepsilon_{i1}\ \varepsilon_{i2}\ \ldots\ \varepsilon_{ip}]^t,\\
\varepsilon_{i\tau} &= j2\pi\left(\Delta\delta_{xi}\left(\frac{m}{MT_x}+\frac{l_x}{T_x}\right)\right.\\
&\quad \left.+\Delta\delta_{yi}\left(\frac{n}{NT_y}+\frac{l_y}{T_y}\right)\right)\phi_{i\tau},\\
\boldsymbol{E} &= [\Delta\boldsymbol{Y}_1\ \Delta\boldsymbol{Y}_2\ \ldots\ \Delta\boldsymbol{Y}_k]^t,\\
\boldsymbol{N} &= [N_1\ N_2\ \ldots\ N_k]^t,\\
\boldsymbol{F} &= [F_{mn}(1)\ F_{mn}(2)\ \ldots\ F_{mn}(p)]^t.
\end{aligned}$$

Furthermore, the superscript t denotes the transpose operator, $p \triangleq 4L_xL_y$, Z_i and N_i are, respectively, the values of the DFTs of the ith noisy undersampled frame and the additive noise at a generic index 2-tuple (m,n) in the wavenumber domain. $F_{mn}(i)$ is the ith interpolated component at index 2-tuple (m,n). This 2-tuple is suppressed in Z_i, N_i, $\boldsymbol{Z}$ and $\boldsymbol{N}$ for brevity. Without loss of generality, N_i is assumed to be zero mean.

In [5, 6], the TLS algorithm [7] developed for the real-valued data is successfully adapted for high resolution image reconstruction using the model given in Equation (3).

3. PERFORMANCE ANALYSIS OF THE TLS ALGORITHM FOR IMAGE RECONSTRUCTION

The TLS theory has been studied extensively in the case of real-valued data [7]. In [5] and [6], the complex-valued data problem has been transformed to an equivalent real-valued data problem to which the TLS theory is applied. In this paper, theoretical analysis of the performance of the TLS algorithm [5, 6] for complex-valued data is proposed. It is shown that the errors originating from the additive noise N_i and the inaccurate estimates of δ_{xi} and δ_{yi} can be suppressed simultaneously by applying the TLS algorithm [5, 6], provided certain constraints are satisfied.

Let C_r and C_i denote, respectively, the real and imaginary parts of a complex number C. Then, (3)

can be rewritten as (4) and (5) below.

$$Z_r = [\ \Phi_r + E_r \quad -\Phi_i - E_i\ ]\begin{pmatrix} F_r \\ F_i \end{pmatrix} + N_r, \quad (4)$$

$$Z_i = [\ \Phi_i + E_i \quad \Phi_r + E_r\ ]\begin{pmatrix} F_r \\ F_i \end{pmatrix} + N_i. \quad (5)$$

The preceding two equations may be combined as

$$U = [A + \Delta A]V + \Delta W, \quad (6)$$

where

$$U = \begin{pmatrix} Z_r \\ Z_i \end{pmatrix}, \quad V = \begin{pmatrix} F_r \\ F_i \end{pmatrix},$$

$$\Delta W = \begin{pmatrix} N_r \\ N_i \end{pmatrix}, \quad A = \begin{pmatrix} \Phi_r & -\Phi_i \\ \Phi_i & \Phi_r \end{pmatrix},$$

$$\Delta A = \begin{pmatrix} E_r & -E_i \\ E_i & E_r \end{pmatrix}.$$

Suppose that there are k noisy undersampled images available. It has been shown [7, p. 242] that

$$\mathrm{cov}(V) \approx (1 + \|V\|^2)\sigma^2 \left(A^t A - k\sigma^2 I\right)^{-1}, \quad (7)$$

if the row vectors of $[\Delta A\ \Delta W]$ are independent and identically distributed with common zero mean vector and common covariance matrix $\sigma^2 I$ where σ^2 (> 0) is a small unknown constant. In (7),

$$A^t A = \begin{pmatrix} \Phi_r^t\Phi_r + \Phi_i^t\Phi_i & \Phi_i^t\Phi_r - \Phi_r^t\Phi_i \\ \Phi_r^t\Phi_i - \Phi_i^t\Phi_r & \Phi_r^t\Phi_r + \Phi_i^t\Phi_i \end{pmatrix}. \quad (8)$$

Equation (8) can be further simplified by making use of (9) and (10) below. As the number of frames k increases, it is not difficult to verify that the Hermitian matrix $\Phi^H\Phi$, where the superscript H denotes Hermitian conjugate, asymptotically approaches a diagonal matrix with nonnegative entries. Therefore, for a large but finite k,

$$\Phi^H\Phi \approx \left[\mathrm{diag}\left(\frac{k}{T_x^2T_y^2} \cdots \frac{k}{T_x^2T_y^2}\right)\right]_{p\times p}. \quad (9)$$

On the other hand,

$$\Phi^H\Phi = \Phi_r^t\Phi_r + \Phi_i^t\Phi_i + j(\Phi_r^t\Phi_i - \Phi_i^t\Phi_r). \quad (10)$$

From(9) and (10), we can have

$$\Phi_r^t\Phi_r + \Phi_i^t\Phi_i \approx \left[\mathrm{diag}\left(\frac{k}{T_x^2T_y^2} \cdots \frac{k}{T_x^2T_y^2}\right)\right]_{p\times p}, \quad (11a)$$

$$\Phi_r^t\Phi_i - \Phi_i^t\Phi_r \approx 0. \quad (11b)$$

Then, from (8), (11a), and (11b), A^tA approximates a $2p \times 2p$ diagonal matrix as below.

$$A^tA \approx \left[\mathrm{diag}\left(\frac{k}{T_x^2T_y^2} \cdots \frac{k}{T_x^2T_y^2}\right)\right]_{2p\times 2p}. \quad (12)$$

By making use of (12), (7) can be rewritten as

$$\mathrm{cov}(V) \approx \frac{1}{k}(1 + \|V\|^2)\frac{\sigma^2T_x^2T_y^2}{1 - \sigma^2T_x^2T_y^2}I. \quad (13)$$

If the additive noise is Gaussian and σ^2 is small, then it has been shown that $\mathrm{cov}(V)$ can reach the Cramér-Rao lower bound [7, p. 243]. In other words, the image reconstructed using the TLS algorithm [5, 6] has minimum variance with respect to all unbiased estimates. As the number of undersampled images, k, increases, the quality of the reconstructed image becomes better and better as shown in (13). In the next section, computer simulation results are provided.

4. COMPUTER SIMULATION RESULTS

In this computer-simulated example, a high resolution 256×256 image, as shown in Figure 1, is recursively reconstructed from a set of low resolution 128×128 noisy frames which are shifted with respect to a reference frame. These shifts or displacements are not accurately known. In order to reduce the size of the arrays to be processed, the input image was partitioned into 16 nonoverlapping sections each of size 32×32 and the recursive TLS algorithm developed in [5, 6] is independently applied to each such section. For each one of these 16 sections, the interpolation problem corresponds to the reconstruction of a 64×64 image from a sequence of shifted low resolution 32×32 noisy input frames when the interframe displacements are not accurately known. To generate the $k = 16$ shifted low resolution input frames required in the simulation from the available data, we use the DFT-based interpolation technique described in [6]. After assigning one of the input frames to be the reference frame, we label it as frame number 1, and the remaining ones are sequentially labeled from $i = 2$ to $i = 16$. The relative shifts of these frames with respect to the frame number 1 along the x and y axes are denoted, respectively, by δ_{xi} and δ_{yi} for $i = 2, 3, \ldots, 16$. The estimation errors, $\Delta\delta_{xi}$ and $\Delta\delta_{yi}$, of δ_{xi} and δ_{yi}, $i = 2, 3, \ldots, 16$, are assumed to be uniformly distributed over $[-\frac{1}{8}, \frac{1}{8}]$. Subsequently, each frame is corrupted by additive noise with SNR level of 20 dB. To illustrate the intermediate steps of the reconstruction, we provide a generic set of sixteen 32×32 frames, as shown in Figure 2, corresponding to a section of the low resolution noisy image. From this set, a

64×64 image is obtained by application of the recursive TLS algorithm. The sequence of estimates leading to the construction of the high resolution filtered image is shown in Figure 3. As expected, the first few iterations of the recursive TLS algorithm provide poor estimates because the algorithm is in a transient stage. Subsequently, the estimates improve.

When the undersampled images are blurred, a post-deblurring approach using Wiener filtering is proposed in [6]. In this computer simulation, the frames are corrupted by a 6-pixel uniform linear motion blur and 30 dB SNR additive noise. Figures 4 and 5 show, respectively, the high resolution images obtained by applying the recursive TLS algorithm without post-deblurring and with post-deblurring.

5. CONCLUDING REMARKS

In computer simulations, the resolution of the reconstructed image increased noticeably when the number of noisy undersampled frames reaches $4L_xL_y$ $(= p)$, the minimum number of undersampled frames needed. As k increases, the quality of the reconstructed image continues to improve. However, the improvement reaches a level of saturation after several frames $k > p$. This may be due to the assumption made to derive (7).

As pointed out in [6], the TLS algorithm outperforms the constrained least squares algorithm when the blurs in the undersampled frames are identical. If the undersampled frames are obtained from the same camera, it is reasonable to assume that the blurs in different frames are identical. The performance analysis proposed here further demonstrates the advantage of using the TLS algorithm in high resolution image reconstruction from noisy undersampled frames when the displacement of any frame with respect to a reference frame is not accurately estimated. Extensive simulations conducted and reported in [6] support the implications of performance analysis reported in this paper. Future research is planned on generalizing the techniques in [5], [6], and this paper to the case of multispectral frames.

Figure 1: Original image - A girl

6. REFERENCES

[1] S. Peleg, D. Kelen, and Schweitzer, "Improving image resolution using subpixel motion," *Pattern Recogition Letters*, vol. 5, pp. 223–226, 1987.

[2] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in *Proc. of IEEE International conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. III, San Francisco, CA, pp. 169–172, 1992.

[3] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP: Graphical Models and Image Processing*, vol. 54, pp. 181–186, March 1992.

[4] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1013–1027, June 1990.

[5] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes," in *Proc. of IEEE International conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. V, Minneapolis, MN, pp. 269–272, 1993.

[6] H. C. Kim, *High Resolution Image Reconstruction from Undersampled Multiframes*. PhD thesis, The Pennsylvania State University, 1994.

[7] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1991.

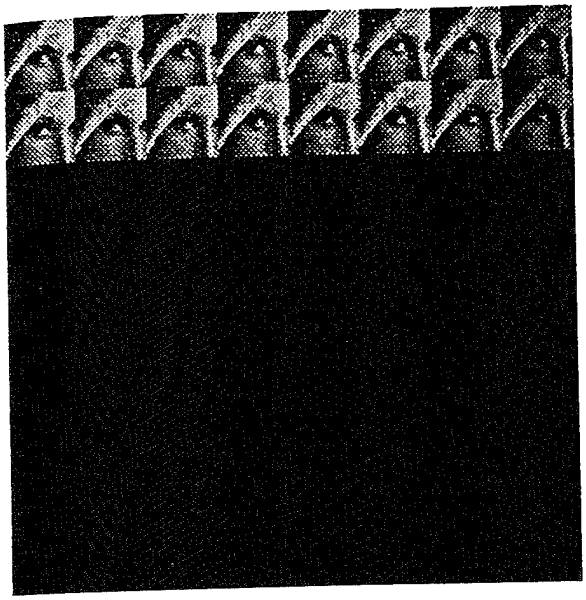


Figure 2: Sixteen 32×32 input girl frames

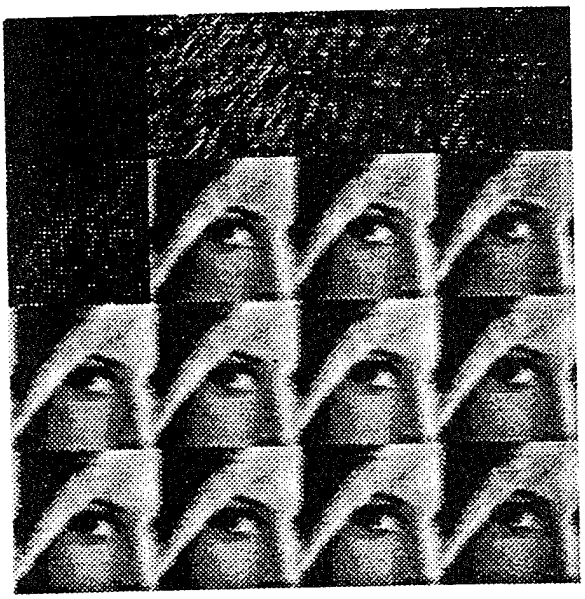


Figure 4: A 64×64 girl image reconstruction sequence without post-deblurring

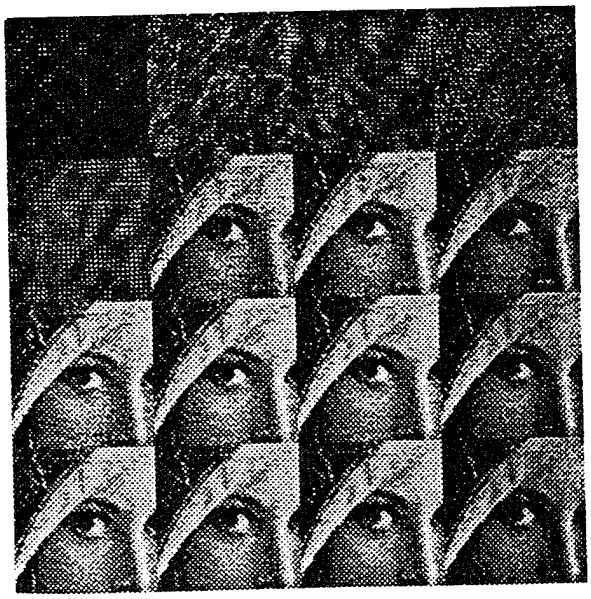


Figure 3: A 64×64 girl image reconstruction sequence

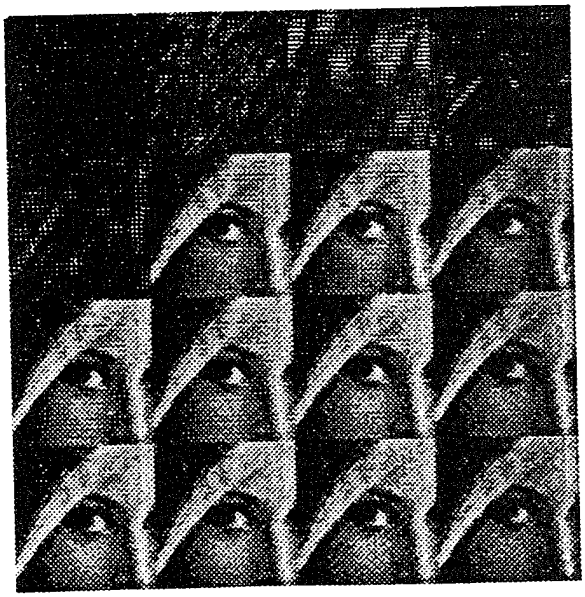


Figure 5: A 64×64 girl image reconstruction sequence with post-deblurring